Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 121 (2011) 1217–1244

stochastic processes and their applications

www.elsevier.com/locate/spa

Spectral estimation of the Lévy density in partially observed affine models

Denis Belomestny

Duisburg-Essen University, Duisburg, Germany

Received 7 December 2009; received in revised form 8 February 2011; accepted 9 February 2011
Available online 24 February 2011

Abstract

The problem of estimating the Lévy density of a partially observed multidimensional affine process from low-frequency and mixed-frequency data is considered. The estimation methodology is based on the log-affine representation of the conditional characteristic function of an affine process and local linear smoothing in time. We derive almost sure uniform rates of convergence for the estimated Lévy density both in mixed-frequency and low-frequency setups and prove that these rates are optimal in the minimax sense. Finally, the performance of the estimation algorithms is illustrated in the case of the Bates stochastic volatility model.

Keywords: Affine processes; Mixed-frequency data; Estimation; Spectral method

1. Introduction

The problem of nonparametric statistical inference for jump processes or more generally for semimartingale models has a long history and goes back to the works of Rubin and Tucker [17] and Basawa and Brockwell [3]. The recent revival of interest in this topic documented, for example, in [10], is mainly related to the wide availability of financial and economical time series data and new types of statistical issues that have not been addressed before. For instance, there is now considerable evidence (see, e.g. [7]) that most financial time series contain a continuous martingale component. This is why in a number of recent works the problem of estimating some characteristics of jumps for the general semimartingale models with a nonzero continuous part was studied. In fact, without any further assumptions such kind of statistical inference would not be possible because the behavior of the jump component becomes statistically indistinguishable

E-mail address: denis.belomestny@uni-due.de.

0304-4149/$ - see front matter
doi:10.1016/j.spa.2011.02.001

from the behavior of the diffusion part as the activity of small jumps increases. In the case of Lévy processes the activity of small jumps can be measured by the so-called Blumenthal–Getoor index. The nearer the Blumenthal–Getoor index to 2, the more difficult becomes the problem of separating jump and diffusion components and hence the problem of statistical inference on the characteristics of jumps (see, e.g. [16]). Suppose that the values of a process $X(t)$ on a time grid $\pi = \{t_0, t_1, \ldots, t_n\}$ are observed. If $|\pi|$ is small (high-frequency data), then a large increment $X(t_i) - X(t_{i-1})$ indicates that a jump occurred between time t_{i-1} and t_i. Based on this insight and the continuous-time observation analogue, inference for various characteristics of jumps of the underlying semimartingale can be conducted. For example, in [1] the problem of statistical inference on the degree of jump activity in the general semimartingale models based on high-frequency data was considered. They proposed an estimation procedure which is able to "see through" the continuous part of the semimartingale and consistently estimate the degree of small jump activity under some restrictions on the structure of the underlying semimartingale. In fact, these restrictions keep the highest degree of activity of small jumps away from 2, thus allowing for a consistent estimation of the degree of jump activity.

In this paper we focus on a special class of semimartingale models, namely the so-called affine models. Affine Itô–Lévy models are nowadays rather popular in financial and econometric modeling. Due to their analytical tractability on the one hand and their rather rich dynamics and implied volatility patterns, on the other hand, they are particularly useful in the context of option pricing. Many well known models such as the Heston and Bates stochastic volatility models fall into the class of affine Itô–Lévy models. Option pricing in these models can be conveniently done via the Fourier method. The literature on affine processes is rather extensive. Let us mention two seminal papers of Duffie et al. [9] and Duffie et al. [8], where theoretical analysis of regular affine models was conducted.

In this work we consider the problem of estimating the characteristics of jumps in a class of affine models with a nonzero continuous part, where it is assumed that only the first few components of the underlying affine process X are observable at low or mixed frequencies. We propose an approach based on the log-affine representation of the conditional characteristic function of an affine process. This representation together with some transformation allows one to consistently estimate the characteristics of the jump component from low-frequency and mixed-frequency data under some prior bound on the highest degree of activity of small jumps. We present uniform convergence rates for the so constructed estimate of the transformed Lévy density which turn out to be optimal in the minimax sense. As the main technical result, that may be of independent interest, we provide exponential inequalities on the probability of large deviations for the kernel type empirical processes in the uniform metric for the case of weakly dependent random variables.

The problem of parametric estimation of the characteristics of an affine jump-diffusion process (processes with finite intensity of jumps) $X(t)$ from high-frequency time series of the asset $S(t) = \exp(X(t))$ has been recently considered in the literature by Singleton [18] and Bates [5]. In Singleton [18] the general method of moments (GMM) based on the empirical characteristic function was employed and the asymptotic properties of the corresponding estimator are investigated. Bates [5] proposed a filtration-based maximum likelihood methodology for estimating the parameters and the realizations of latent affine processes. Since the characteristics of a general affine process are a priori an infinite-dimensional object, any parametric approach is always exposed to the problem of misspecification, especially if there is no inherent economic foundation for the parameters and they are only used to generate different shapes of possible jump distributions. The problem of semi-parametric inference for the characteristics of special type affine processes $X(t)$ was studied in the literature as well. In the case of high-frequency observations, the problem of nonparametric inference on the Lévy measure of the time-changed Lévy processes, belonging sometimes to the class of affine processes, has been recently studied in [11]. In Jongbloed et al. [12] the case of a one-dimensional Lévy driven Ornstein–Uhlenbeck process, an affine process with a zero diffusion part, was considered. The authors assumed that the corresponding jump component is self-decomposable and proposed a cumulant M-estimator to estimate the so-called canonical function of the driving self-decomposable process from low-frequency data. As to the special case of Lévy processes, semi-parametric estimation for pure Lévy models under low-frequency data has recently been studied in [16]. Let us mention that in [16] the diffusion component is assumed to be known. Thus, all the above works do not encounter the problem of separating diffusion and jump components as the activity of small jumps increases. Furthermore, the challenge of devising nonparametric estimation methods for the Lévy density in general affine models lies in the fact that the structure of the conditional characteristic function does not have such an explicit form as in the case of pure Lévy processes and is related to the parameters of the underlying affine process not directly but via a Riccati equation. Last but not the least: the increments of the general affine process are not independent, hence advanced tools from the time series analysis have to be used.

The paper is organized as follows. In Section 2 we introduce the main object of our study, the affine Itô–Lévy processes and formulate the main existence result. In Section 3 the main ideas behind our estimation methodology are sketched and the notation is introduced. The estimation algorithms for the cases of mixed-frequency and low-frequency data are presented in Section 4 and Section 5, respectively. The asymptotic properties of the constructed estimates are studied in Section 6. Section 7 contains some numerical examples. The exponential inequalities for the kernel type empirical processes are given in Section 8. Finally, the proofs of the main results are collected in Section 9.

2. Main setup

Let us fix a probability space $(\Omega, \mathcal{F}, P)$ and an information filtration $(\mathcal{F}_t)_{t\geq 0}$. The process $X(t)$ is an affine process if it is a stochastically continuous, time-homogeneous Markov process with the state space $\mathcal{D} \subset \mathbb{R}^d$, such that the conditional characteristic function of $X(t)$ given $X(0)$ is an affine function of the initial state $X(0)$:

$$\phi(u|s, x) := \mathrm{E}(\mathrm{e}^{\mathrm{i}u^\top X(s)}|X(0) = x) = \mathrm{e}^{\psi_0(u,s)+x^\top\psi_1(u,s)}, \quad u \in \mathbb{R}^d. \tag{1}$$

The affine process $X(t)$ is called regular, if the derivatives

$$F_0(u) := \left.\frac{\partial\psi_0(u, s)}{\partial s}\right|_{s=0}, \qquad F_1(u) := \left.\frac{\partial\psi_1(u, s)}{\partial s}\right|_{s=0}$$

exist and are continuous at $u = 0$. As was recently shown by Keller-Ressel et al. [13], any affine process is, in fact, regular. The following theorem provides the characterization of affine processes and is proved in [8].

Theorem 2.1. *If $(X(t))_{t\geq 0}$ is a regular affine process, then the complex valued functions ψ_0 and ψ_1 satisfy the following (generalized) Riccati equations*

$$\frac{\partial\psi_0(u, s)}{\partial s} = F_0(\psi_1(u, s)), \quad \psi_0(u, 0) = 0, \tag{2}$$

$$\frac{\partial \psi_1(u,s)}{\partial s} = F_1(\psi_1(u,s)), \quad \psi_1(u,0) = \mathrm{i}u, \tag{3}$$

where

$$F_0(z) = (\alpha^{(0)}z, z) + (z, \beta^{(0)}) - \gamma^{(0)} + \int_{\mathcal{D}\setminus\{0\}} (\mathrm{e}^{z^\top u} - 1 - (\chi(u), z))\nu^{(0)}(\mathrm{d}u),$$

$$F_{1,j}(z) = (\alpha_j^{(1)}z, z) + (z, \beta_j^{(1)}) - \gamma_j^{(1)} + \int_{\mathcal{D}\setminus\{0\}} (\mathrm{e}^{z^\top u} - 1 - (\chi(u), z))\nu_j^{(1)}(\mathrm{d}u)$$

for $j = 1, \ldots, d$ *and* $\chi(u) = (\chi_1(u), \ldots, \chi_d(u))$ *with*

$$\chi_k(u) = \begin{cases} 0, & u_k = 0, \\ (1 \wedge |u_k|)\dfrac{u_k}{|u_k|}, & \text{otherwise} \end{cases}$$

for $k = 1, \ldots, d$. *Here* $\alpha = (\alpha^{(0)}, \alpha^{(1)}) \in \mathbb{R}^{d\times d} \times \mathbb{R}^{d\times d\times d}$, $\beta = (\beta^{(0)}, \beta^{(1)}) \in \mathbb{R}^d \times \mathbb{R}^{d\times d}$, $\gamma = (\gamma^{(0)}, \gamma^{(1)}) \in \mathbb{R} \times \mathbb{R}^d$ *and* $\nu = (\nu^{(0)}, \nu_1^{(1)}, \ldots, \nu_d^{(1)})$ *is a vector of measures on* $\mathbb{R}^d$, *satisfying*

$$\int_{\mathcal{D}\setminus\{0\}} \|\chi(u)\|_2^2 \nu^{(0)}(\mathrm{d}u) < \infty, \qquad \int_{\mathcal{D}\setminus\{0\}} \|\chi(u)\|_2^2 \nu_j^{(1)}(\mathrm{d}u) < \infty, \quad j = 1, \ldots, d,$$

where here and in the sequel $\|x\|_2 := \sqrt{x_1^2 + \cdots + x_d^2}$ *for any* $x \in \mathbb{R}^d$.

Under some admissibility conditions a regular affine process $X(t)$ is a Feller process in the domain $\mathcal{D} = \mathbb{R}^m \times \mathbb{R}_+^{d-m}$ (see [8, Section 2]), where the function F_0 corresponds to the state-independent part of the infinitesimal operator and F_1 is related to the state-dependent one. The admissibility conditions imply, in particular, that

$$\alpha_{ij}^{(0)} = 0, \quad \text{if } i \in \{m+1, \ldots, d\} \text{ or } j \in \{m+1, \ldots, d\} \tag{4}$$

and $\alpha_j^{(1)} \equiv 0$ for $j = 1, \ldots, m$. In the sequel we assume that the above admissibility conditions hold. Moreover, we restrict our analysis to the class of affine processes with state-independent jumps, i.e.,

$$\nu_1^{(1)} \equiv \ldots \equiv \nu_d^{(1)} \equiv 0. \tag{5}$$

On the one hand, this assumption reduces the dimensionality of the jump component of X. On the other hand, the class of affine models satisfying (5), remains rather large and includes such well known models as the Heston, Bates and Barndorff–Nielsen and Shephard stochastic volatility models. In this paper we study the problem of statistical inference based on partially observed affine processes. In particular, we assume that only the first m components of the process X are observed (as it is usual in the case of stochastic volatility models). As a result, in order to ensure identifiability, we have to assume additionally that $\operatorname{supp} \nu^{(0)} \subset \mathbb{R}^m$, i.e., the positive part of the process X does not have jumps.

3. Main ideas

Assume that the process $X(t)$ is stationary with the stationary distribution π. Fix some $x \in \mathbb{R}^m$, $s \in \mathbb{R}_+$ and denote

$$\begin{aligned}\psi(v|s,x) = {} & \psi_0((v, 0, \ldots, 0), s) + (x, \mathrm{E}_\pi[X_{m+1}(0)], \ldots, \mathrm{E}_\pi[X_d(0)])^\top \\ & \times \psi_1((v, 0, \ldots, 0), s)\end{aligned} \tag{6}$$

for any $v \in \mathbb{R}^m$. Introduce the function

$$\Psi(v|s,x) := \int_{[-1,1]^m} (\psi(v|s,x) - \psi(v+w|s,x))\,\mathrm{d}w. \tag{7}$$

Let us now investigate the behavior of the function Ψ as $s \to 0$ ("short time asymptotic") and $s \to \infty$ ("long time asymptotic").

Short time asymptotic. Due to (2) and (3), it holds

$$\Psi_0(v) = \left.\frac{\partial \Psi(v|s,x)}{\partial s}\right|_{s=0} = \mathcal{L} + \int_{\mathbb{R}^m} \mathrm{e}^{\mathrm{i}z^\top v} \rho^{(0)}(\mathrm{d}z) \tag{8}$$

for some $\mathcal{L} \in \mathbb{R}$ depending linearly on x, where

$$\rho^{(0)}(\mathrm{d}z) := 2^m \prod_{k=1}^m \left(1 - \frac{\sin z_k}{z_k}\right) \nu^{(0)}(\mathrm{d}z) \tag{9}$$

is a finite measure on $\mathbb{R}^m$.

Long time asymptotic. If the maximal eigenvalue of the matrix $(\beta_{ij}^{(1)})_{1\le i,j\le d}$ is negative, then the affine process X is ergodic, possesses a stationary distribution π (see [14]) and it holds $\phi(v,0,\dots,0|s,x) \to \phi_\pi(v)$ as $s \to +\infty$, with

$$\phi_\pi(v) = \exp\left(\int_0^\infty F_0(\psi_1((v,0\dots,0),s))\,\mathrm{d}s\right), \quad v \in \mathbb{R}^m.$$

If

$$\beta_{m+1}^{(0)} = \cdots = \beta_d^{(0)} = 0, \tag{10}$$

then the admissibility condition (4) implies

$$\begin{aligned}\psi_\pi(v) &= \log\{\phi_\pi(v)\} = -(v, \widetilde{\alpha}^{(0)} v) + \mathrm{i}(\widetilde{\beta}^{(0)}, v)\\ &\quad + \int_0^\infty \int_{\mathbb{R}^m} (\mathrm{e}^{\mathrm{i}(\mathrm{e}^{s\mathfrak{B}} v, w)} - 1 - \mathrm{i}(\chi(w), \mathrm{e}^{s\mathfrak{B}} v))\nu^{(0)}(\mathrm{d}w)\,\mathrm{d}s\end{aligned} \tag{11}$$

for some $\widetilde{\alpha}^{(0)} \in \mathbb{R}^{m\times m}$, $\widetilde{\beta}^{(0)} \in \mathbb{R}^m$ and $\mathfrak{B} = (\beta_{ij}^{(1)})_{1\le i,j\le m}$. Therefore

$$\Psi_\pi(v) = \lim_{s\to+\infty} \Psi(v|s,x) = \widetilde{\mathcal{L}} + \int_{\mathbb{R}^m} \mathrm{e}^{\mathrm{i}z^\top v} \widetilde{\rho}^{(0)}(\mathrm{d}z) \tag{12}$$

with some $\widetilde{\mathcal{L}} > 0$ and

$$\widetilde{\rho}^{(0)}(\mathrm{d}z) := 2^m \prod_{k=1}^m \left(1 - \frac{\sin z_k}{z_k}\right) \widetilde{\nu}^{(0)}(\mathrm{d}z), \tag{13}$$

where for any set $A \subset \mathcal{B}(\mathbb{R}^m)$

$$\widetilde{\nu}^{(0)}(A) = \int_0^\infty \int_{\exp(s\mathfrak{B}^*)q\in A} \nu^{(0)}(\mathrm{d}q)\,\mathrm{d}s.$$

In the sequel we shall assume that the measure $\nu^{(0)}$ is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^m$ and has a bounded density. Then $\widetilde{\nu}^{(0)}$ also has a density given by

$$\widetilde{\nu}^{(0)}(z) = \int_0^\infty |\mathrm{e}^{-s\mathfrak{B}^*}|\nu^{(0)}(\mathrm{e}^{-s\mathfrak{B}^*}z)\,\mathrm{d}s$$

and the admissibility conditions imply that

$$\max_{k=1,\dots,m} \inf\left\{ r \geq 0 : \int_{\{|x_1|>1,\dots,|x_{k-1}|>1,\,|x_k|\leq 1,\,|x_{k+1}|>1,\dots,|x_m|>1\}} |x_k|^r \nu^{(0)}(\mathrm{d}x) < \infty \right\} < 1.$$

Hence, the measures $\rho^{(0)}$ and $\widetilde{\rho}^{(0)}$ are absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^m$ as well and possess bounded densities denoted (with some abuse of notation) by $\rho^{(0)}(x)$ and $\widetilde{\rho}^{(0)}(x)$, respectively. Functions $\Psi_0(v)$ and $\Psi_\pi(v)$ satisfy, due to the Riemann–Lebesgue theorem, the following asymptotic relations

$$\lim_{\|v\|\to\infty} \Psi_0(v) = \mathcal{L},$$
$$\lim_{\|v\|\to\infty} \Psi_\pi(v) = \widetilde{\mathcal{L}}.$$

The above relations together with (8) and (11) indicate that one can estimate the Fourier transforms of $\rho^{(0)}(x)$ and $\widetilde{\rho}^{(0)}(x)$, if some estimates for the functions $\Psi_0(u)$ and $\Psi_\pi(u)$ are available. In order to estimate Ψ_0 using formula (8) we need to perform numerical differentiation of the function $\Psi(v|s,x)$ in s. This calls for the "high-frequency" data. On the other hand, in order to estimate the expectation in Ψ by a kind of averaging, we need an ergodic theorem which usually holds only under low-frequency sampling. It turns out that by mixing low- and high-frequency data one can consistently estimate $\Psi_0(u)$. As to the function Ψ_π, it can be estimated from low-frequency data.

Remark 3.1. In the case of the more general affine processes, i.e., in the case where $\nu_k^{(1)} \not\equiv 0$ for some $k \in \{1,\dots,m\}$, one can use a similar strategy to reconstruct $\nu_k^{(1)}$, $k = 1,\dots,m$. Indeed, in the general case the function Ψ_0 takes the form

$$\Psi_0(v) = \mathcal{L} + \int_{\mathbb{R}^m} \mathrm{e}^{\mathrm{i}z^\top v}\rho^{(0)}(\mathrm{d}z) + \sum_{j=1}^m x_j \int_{\mathbb{R}^m} \mathrm{e}^{\mathrm{i}z^\top v}\rho_j^{(1)}(\mathrm{d}z), \tag{14}$$

where each measure $\rho_j^{(1)}(\mathrm{d}z)$ is related to $\nu_j^{(1)}(\mathrm{d}z)$ in the same way as $\rho^{(0)}(\mathrm{d}z)$ was related to $\nu^{(0)}(\mathrm{d}z)$. Therefore, one can reconstruct the Fourier transforms of all measures $\rho_j^{(1)}(\mathrm{d}z)$, $j = 1,\dots,m$, if one is able to recover function $\Psi(u|s,x)$ and its derivative in s for (at most) m linearly independent vectors x. In principle, the approach presented in the next section allows one to estimate $\Psi(v|s,x)$ for an arbitrary number of vectors x.

4. Estimation of $\rho^{(0)}(z)$ in the mixed-frequency setup

4.1. Observations

For our theoretical study we adopt the observational model based on the mixed-frequency random sampling. In particular, we assume that a trajectory of a partially observed process X

containing the pairs

$$(X^{(m)}(t_j), X^{(m)}(t_j + \delta_j)), \quad j = 1, \ldots, n,$$

is observable, where $X^{(m)}(t) = (X_1(t), \ldots, X_m(t)), \delta_j, j = 1, \ldots, n$, are i.i.d. random variables on $[0, T]$ for some fixed $T > 0$ with a common density $p_\delta(x)$ and $\min_j(t_{j+1} - t_j) > T$. The latter assumption implies that the time horizon $T + t_n$ of observations tends to infinity as $n \to \infty$. In the sequel (see Assumption (AG) in Section 6 and Remark 6.2) we will additionally assume that the density of the r.v. (δ_j) does not vanish in the vicinity of 0, meaning that the r.v. (δ_j) can, with positive probability, take values that are arbitrary close to 0. This is the reason why we call the above observation's scheme a mixed-frequency setup. Mixing low-frequency and high-frequency data has become a rather popular technique in recent years. It has been, for example, used to improve the volatility estimation (see, e.g., [20]) or to achieve better forecasts in macro-economical models (see, e.g. [2] and references therein). Let us also note that the high-frequency data is usually not equidistant: the observations are more frequent for busy trading days. In such a situation our random sampling scheme may be appropriate. As we will see, the above sampling scheme will also allow us to consistently estimate the function Ψ_0 in (8). Indeed, while the high-frequency sampling scheme (small values of (δ_j)) makes it possible to estimate the derivative of the function $\Psi(v|s, x)$ in s, the low-frequency data allows us to consistently estimate the function $\Psi(v|s, x)$ by its empirical counterpart. The condition $t_{j+1} - t_j - \delta_j > 0$ ensures that the subsequent pairs are only weakly dependent and a kind of ergodic theorem can be applied.

4.2. *Estimation of* $\Psi_0(v)$

Assume again that the process $X(t)$ is stationary with the stationary distribution π. In this section we shall, for any fixed $x \in \mathbb{R}^m$, estimate the quantity

$$\partial\phi((v, 0 \ldots, 0)|s, (x, \mathrm{E}_\pi[X_{m+1}(0)], \ldots, \mathrm{E}_\pi[X_d(0)]))\partial s|_{s=0}, \tag{15}$$

by local polynomial smoothing (local polynomial of degree l in x and local linear in s) of the empirical characteristic process

$$Z_j(v) = \exp(\mathrm{i} v^\top X^{(m)}(t_j + \delta_j)), \quad v \in \mathbb{R}^m, \quad j = 1, \ldots, n.$$

Since the process $X(t)$ is stationary, (15) is equal to

$$\mathrm{E}_{\pi(X_{m+1},\ldots,X_d)}[\partial\phi((v, 0 \ldots, 0)|s, (x, X_{m+1}, \ldots, X_d))/\partial s|_{s=0}].$$

The latter quantity can be estimated by performing the local polynomial smoothing w.r.t. the first m components of the process $X(t)$ and averaging w.r.t. to the conditional distribution of the remaining coordinate processes $X_{m+1}(t), \ldots, X_d(t)$. This is basically what we do next. Fix some $x \in \mathbb{R}^m$. For some $h_1 > 0, h_2 > 0$, an integer $l \geq 0$ and a function $K : \mathbb{R}^{1+m} \to \mathbb{R}_+$, let $(\mathcal{Q}_{0,n}, \mathcal{Q}_{1,n})$ be a solution of the following optimization problem

$$\min_{(\mathcal{Q}_0, \mathcal{Q}_1)} \left\{ \sum_{j=1}^{n} w_j [Z_j(v) - \mathcal{Q}_0(X^{(m)}(t_j) - x) - \delta_j \mathcal{Q}_1(X^{(m)}(t_j) - x)]^2 \right\} \tag{16}$$

with $w_j = K(\delta_j/h_1, (X^{(m)}(t_j) - x)/h_2)$, where the minimization is performed over the set of all polynomials $\mathcal{Q}_0$ and $\mathcal{Q}_1$ on $\mathbb{R}^m$ of degree l. Now define the local polynomial estimates for

$\exp(\psi(v|0,x))$ and $\partial_s\psi(v|0,x)$, where $\psi(v|0,x)$ is defined in (6), by

$$\phi_n(v) = \mathcal{Q}_{0,n}(0), \qquad \phi_{s,n}(v) = \mathcal{Q}_{1,n}(0),$$

respectively. Furthermore, define an estimate for $\Psi_0(v)$ by plugging the estimate $\phi_{s,n}$ into (7):

$$\Psi_{0,n}(v) := \int_{[-1,1]^m} (\phi_{s,n}(v) - \phi_{s,n}(v+w))\,\mathrm{d}w.$$

Let $\pi_{0,\mathbf{m}}$ and $\pi_{1,\mathbf{m}}$ denote the coefficients of the polynomials $\mathcal{Q}_{0,n}$ and $\mathcal{Q}_{1,n}$, respectively, indexed by the multi-index $\mathbf{m} \in \mathbb{N}^m$, i.e., $\mathcal{Q}_{k,n}(z) = \sum_{|\mathbf{m}|\le l} \pi_{k,\mathbf{m}} z^{\mathbf{m}}$, $k = 0, 1$. Introduce the vectors $\Pi_k = (\pi_{k,\mathbf{m}})_{|\mathbf{m}|\le l}$ and $S_k = (S_{k,\mathbf{m}})_{|\mathbf{m}|\le l}$, $k = 0, 1$, with

$$S_{k,\mathbf{m}} = \frac{1}{nh_1h_2^m}\sum_{j=1}^n Z_j(v)\left(\frac{\delta_j}{h_1}\right)^k \left(\frac{X^{(m)}(t_j)-x}{h_2}\right)^{\mathbf{m}} w_j.$$

Let $P(z) = (z^{\mathbf{m}})_{|\mathbf{m}|\le l}$ be the vector of all monomials in $\mathbb{R}^m$ of order less than or equal to l and the matrices $\Gamma_k = (\Gamma_{k,\mathbf{m}_1,\mathbf{m}_2})_{|\mathbf{m}_1|,|\mathbf{m}_2|\le l}$, $k = 0, 1, 2$, be defined as

$$\Gamma_{k,\mathbf{m}_1,\mathbf{m}_2} = \frac{1}{nh_1h_2^m}\sum_{j=1}^n \left(\frac{\delta_j}{h_1}\right)^k \left(\frac{X^{(m)}(t_j)-x}{h_2}\right)^{\mathbf{m}_1+\mathbf{m}_2} w_j. \tag{17}$$

Consider now the vector $S = (S_0, S_1)^\top$ and the matrix

$$\Gamma = \begin{pmatrix} \Gamma_0 & \Gamma_1 \\ \Gamma_1 & \Gamma_2 \end{pmatrix}.$$

The following proposition holds.

Proposition 4.1. *If the matrix Γ is positive definite, then there exist unique polynomials $\mathcal{Q}_{0,n}$ and $\mathcal{Q}_{1,n}$ on $\mathbb{R}^m$ of degree l solving (16). Their vectors of coefficients are given by $\Pi = (\Pi_0, \Pi_1) = \Gamma^{-1}S$. As a result*

$$(\phi_n(v), \phi_{s,n}(v))^\top = (P(0), P(0))^\top \Gamma^{-1} S. \tag{18}$$

Proposition 4.1 implies that $\phi_{s,n}(v) = \sum_{j=1}^n \omega_j Z_j(v)$ with some weights ω_j, $j = 1, \ldots, n$. Hence, the following representation for $\Psi_{0,n}$ holds

$$\Psi_{0,n}(v) = \sum_{j=1}^n \bar{\omega}_j Z_j(v), \qquad v \in \mathbb{R}^m,$$

where $\bar{\omega}_j = \omega_j \int_{[-1,1]^m} (1 - \exp(\mathrm{i}v^\top X^{(m)}(t_j)))\,\mathrm{d}v$.

4.3. *Estimation of $\rho^{(0)}(z)$*

Let $\mathcal{K}_{\mathcal{L}}(u)$ be a regularizing kernel supported on $[-1,1]^m$ and let U_n be a sequence of positive numbers tending to ∞. Define an estimate for the limit $\mathcal{L}$ in (8) as

$$\mathcal{L}_n = \int_{\mathbb{R}^m} U_n^{-m} \mathcal{K}_{\mathcal{L}}(v/U_n)\,\Psi_{0,n}(v)\,\mathrm{d}u.$$

Next, we reconstruct the Lévy density $\rho^{(0)}(z)$ using the regularized Fourier inversion formula

$$\rho_n^{(0)}(z) = \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} \mathrm{e}^{-\mathrm{i}v^\top x} \mathcal{K}_\rho(v/U_n)[\Psi_{0,n}(v) - \mathcal{L}_n]\,\mathrm{d}u,$$

where $\mathcal{K}_\rho$ is another regularizing kernel.

5. Estimation of $\widetilde{\rho}^{(0)}$ using low-frequency data

5.1. Observations

We assume that the time series $X^{(m)}(t_1), \ldots, X^{(m)}(t_n)$ is observed, where $t_k,\ k = 1, \ldots, n$, is a deterministic sequence of positive numbers with $t_k - t_{k-1} > \Delta$, $k = 1, \ldots, n$, for some $\Delta > 0$.

5.2. Estimation of Ψ_π

By the Birkhoff ergodic theorem it holds for any $v \in \mathbb{R}^m$

$$\phi_{\pi,n}(v) = \frac{1}{n} \sum_{k=1}^{n} \exp(v^\top X^{(m)}(t_k)) \to \phi_\pi(v), \quad n \to \infty,$$

where $\phi_\pi(v)$ stands for the c.f. of the first m components of $X(t)$ under π. Therefore it is natural to estimate Ψ_π via

$$\Psi_{\pi,n}(v) := \int_{[-1,1]^m} [\log\{\phi_{\pi,n}(v)\} - \log\{\phi_{\pi,n}(v + w)\}]\,\mathrm{d}w.$$

5.3. Estimation of $\widetilde{\rho}^{(0)}$

We again first estimate the limit $\widetilde{\mathcal{L}}$ by

$$\widetilde{\mathcal{L}}_n = \int_{\mathbb{R}^m} U_n^{-m} \mathcal{K}_{\mathcal{L}}(v/U_n)\, \Psi_{\pi,n}(v)\,\mathrm{d}u.$$

Then the transformed density $\widetilde{\rho}^{(0)}(z)$ can be reconstructed using the regularized Fourier inversion formula

$$\widetilde{\rho}_n^{(0)}(z) = \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} \mathrm{e}^{-\mathrm{i}v^\top z} \mathcal{K}_\rho(v/U_n)[\Psi_{\pi,n}(v) - \widetilde{\mathcal{L}}_n]\,\mathrm{d}v.$$

6. Asymptotic analysis

In this section we study the asymptotic properties of the estimates $\rho^{(0)}$ and $\widetilde{\rho}^{(0)}$.

6.1. Assumptions

We need the following assumptions.

(AX) The sequence $X(t_k)$, $k \in \mathbb{N}$, is strongly mixing with the mixing coefficients α_X satisfying

$$\alpha_X(k) \le \bar{\alpha}_0 \exp(-\bar{\alpha}_1 k), \quad k \ge 1,$$

for some $\bar{\alpha}_0 > 0$ and $\bar{\alpha}_1 > 0$.

(AN) The Lévy measure $\nu^{(0)}$ satisfies for some $p > 2$

$$\int_{\{\|x\|>1\}} \|x\|^p \nu^{(0)}(\mathrm{d}x) < \infty,$$

where here and in the sequel $\|\cdot\|$ stands for l_∞ norm.

Let $p_{m,\pi}$ be the density of $X^{(m)}(t)$ under π. For any fixed s and x, consider a matrix

$$\bar{\Gamma} = \begin{pmatrix} \bar{\Gamma}_0 & \bar{\Gamma}_1 \\ \bar{\Gamma}_1 & \bar{\Gamma}_2 \end{pmatrix},$$

where $\bar{\Gamma}_k = (\bar{\Gamma}_{k,\mathbf{m}_1,\mathbf{m}_2})_{|\mathbf{m}_1|,|\mathbf{m}_2|\le l}$ are matrices with the elements

$$\bar{\Gamma}_{k,\mathbf{m}_1,\mathbf{m}_2} = \int_{\mathbb{R}^m} \int_0^\infty t^k z^{\mathbf{m}_1+\mathbf{m}_2} K(t,z) p_\delta(h_1 t) p_{m,\pi}(x + h_2 z)\, \mathrm{d}t\, \mathrm{d}z.$$

Note that $\mathrm{E}\,\Gamma = \bar{\Gamma}$. We make the following assumption about $\bar{\Gamma}$.

(AG) The minimal eigenvalue of the matrix $\bar{\Gamma}$ is bounded away from zero, i.e.,

$$\min_{\|W\|=1} [W^\top \bar{\Gamma} W] \ge \gamma_0$$

with some $\gamma_0 > 0$.

(AP) The density $p_{m,\pi}$ is uniformly bounded on $\mathbb{R}^m$.

(AK1) Kernel K is bounded and is supported on $[0,1] \times [-1,1]^m$.

(AK2) The regularizing kernels $\mathcal{K}_\mathcal{L}$ and $\mathcal{K}_\rho$ are uniformly bounded, are supported on $[-1,1]^m$, integrate to 1 and satisfy

$$\mathcal{K}_\mathcal{L}(u) = 0, \quad \mathcal{K}_\rho(u) = 1, \quad u \in [-a_K, a_K]^m$$

with some $0 < a_K < 1$.

Discussion. Exponentially strongly mixing holds for a wide class of Itô–Lévy processes. In Masuda [14] conditions are formulated that ensure that a multidimensional Itô–Lévy process is exponentially β-mixing and hence exponentially α-mixing.

Example 6.1. Let A be $d \times d$ matrix whose eigenvalues have positive real parts, and let Z be a nontrivial d-dimensional Lévy process. Consider a d-dimensional Ornstein–Uhlenbeck process X given by

$$\mathrm{d}X(t) = -AX(t)\mathrm{d}t + \mathrm{d}Z(t), \quad X_0 \sim \eta.$$

The process X is obviously an affine process satisfying (5). If the Lévy measure ν of the process Z satisfies $\int_{\|x\|>1} \|x\|^q \nu(\mathrm{d}x) < \infty$ and $\int_{\|x\|>1} \|x\|^q \eta(\mathrm{d}x) < \infty$ for some $q > 0$, then X is exponentially β-mixing (see [14, Theorem 2.6]).

Suppose now that X is an affine process with the characteristics χ satisfying admissibility conditions, the condition (5) and $\int_{\|x\|>1} \|x\|^q \nu^{(0)}(\mathrm{d}x) < \infty$ for some $q > 0$. If the maximal eigenvalue of the matrix $\beta^{(1)}$ is negative, then both sequences $X(t_n)$ and $X(t_n + \delta_n)$ are exponentially β-mixing and hence ergodic.

Remark 6.2. Let us remark on the assumption (AG). If for any $R > 0$ the joint density $p_\delta(t) p_{m,\pi}(x)$ is strictly positive on $(0,T] \times \mathcal{B}_R$, where $\mathcal{B}_R$ is a ball of radius R in $\mathbb{R}^m$, then (AG) holds. Suppose, for simplicity, that $p_{m,\pi}(x)$ is supported on $\mathcal{B}_R$ for some $R > 0$

(otherwise a truncation argument combined with an assumption on the tails of $p_{m,\pi}$ can be used) and consider the kernel

$$K(t,z) := \frac{\Gamma(1+(m+1)/2)}{\pi^{(m+1)/2}} \mathbf{1}_{\{\sqrt{t^2+\|z\|^2}\le 1\}}.$$

We have for any $W = (W_1, W_2) \in \mathbb{R}^D \times \mathbb{R}^D$ with $D = m(m+2)\cdot\cdots\cdot(m+l-1)/l!$

$$\begin{aligned} W^\top \bar{\Gamma}(x) W &= \int_0^T \int_{\mathbb{R}^m} \left(\sum_{|\alpha|\le l} W_1^\alpha z_\alpha + t \sum_{|\alpha|\le l} W_2^\alpha z_\alpha \right)^2 \\ &\quad \times K(t,z) p_\delta(h_1 t) p_{m,\pi}(x + h_2 z)\, \mathrm{d}t\, \mathrm{d}z \\ &\ge B \int_{\mathcal{S}(x,R,T)} \left(\sum_{|\alpha|\le l} W_1^\alpha z_\alpha + t \sum_{|\alpha|\le l} W_2^\alpha z_\alpha \right)^2 \mathrm{d}t\, \mathrm{d}z \end{aligned}$$

with some positive constant B and

$$\begin{aligned} \mathcal{S}(x,R,T) := \{&(t,z) \in \mathbb{R}_+ \times \mathbb{R}^m : t^2 + \|z\|^2 \le 1, \\ &\|x + h_2 z\|^2 + h_1^2 t^2 \le R^2 + T^2\}. \end{aligned}$$

Using now the fact that the Lebesgue measure of the set $\mathcal{S}(x,R,T)$ is larger than some positive number λ for all $x \in \mathcal{B}_R$, where λ depends on R, T and d but does not depend on h_1 and h_2, we get

$$\begin{aligned} \inf_{x\in\mathcal{B}_R} [W^\top \bar{\Gamma}(x) W] &\ge B \inf_{\|W\|=1} \inf_{\mathcal{S}:|\mathcal{S}|>\lambda} \int_{\mathcal{S}} \left(\sum_{|\alpha|\le l} W_1^\alpha z_\alpha + t \sum_{|\alpha|\le l} W_2^\alpha z_\alpha \right)^2 \mathrm{d}t\, \mathrm{d}z \\ &\ge \gamma_0 \end{aligned}$$

with some positive γ_0 by the compactness argument.

6.2. *Minimax upper bounds for $\rho_n^{(0)}$*

First, introduce a class of Lévy densities for which we are going to derive the minimax rates of convergence. For any $1 \le \varkappa < 2$ let $\mathfrak{L}_\varkappa$ stand for a class of Lévy densities $\nu^{(0)}$ satisfying

$$\left(\prod_{k=1}^d v_k \right)^\varkappa \mathcal{F}[\rho^{(0)}](v) \le C, \quad v \in \mathbb{R}^m \tag{19}$$

for some $C > 0$, where the function $\rho^{(0)}$ is related to the Lévy density $\nu^{(0)}$ via (9). Here and in the sequel $\mathcal{F}[\rho](u)$ stands for the Fourier transform of a measure ρ.

Remark 6.3. It can be shown that if the Lévy measure $\nu^{(0)}$ satisfies

$$\begin{aligned} &\max_{k=1,\ldots,m} \inf \left\{ r \ge 0 : \int_{\{|x_1|>1,\ldots,|x_{k-1}|>1,\ |x_k|\le 1,\ |x_{k+1}|>1,\ldots,|x_m|>1\}} |x_k|^r \nu^{(0)}(\mathrm{d}x) < \infty \right\} \\ &\quad < 2 - \varkappa, \end{aligned}$$

i.e., the degree of jump activity (for the definition see [1]) of the each component of the process $X(t)$ is less than $2-\varkappa$, then the inequality (23) holds with some $C > 0$.

The first result of this section concerns the asymptotic properties of the estimate $\phi_{s,n}(v)$ constructed in Section 4.2.

Theorem 6.4. *Suppose that the assumptions* (AX), (AN), (AG), (AP) *and* (AK1) *hold. Let $\phi_{n,s}(v)$ be the local polynomial estimate of degree l (in x) for the function $\phi_s(v) = \partial_s \psi(v|0,x)$, where $\psi(v|0,x)$ is defined in (6). Furthermore, let w be a monotone positive Lipschitz function on $\mathbb{R}_+$ such that*

$$0 < w(z) \leq 1/(1+z)^{2(l+1)}, \quad z \in \mathbb{R}_+. \tag{20}$$

Then under choices $h_1 = (n^{-1}\log^{1+r} n)^{(l+1)/(2m+5(l+1))}$ and $h_2 = (n^{-1}\log^{1+r} n)^{1/(m+5(l+1)/2)}$ with some $r > 0$,

$$\mathrm{P}(\sup_{v\in\mathbb{R}^m}[w(\|v\|)|\phi_{n,s}(v) - \phi_s(v)|] > A\zeta_n) \leq Bn^{-1-\delta}, \tag{21}$$

where

$$\zeta_n = [n^{-1}\log^{(1+r)} n]^{\frac{1+l}{m+5(1+l)}} \tag{22}$$

and δ, A and B are some positive constants.

Remark 6.5. The condition (20) on the decay of the weighting function w cannot be, in general, weakened. For example, in the case of a one-dimensional Brownian motion with volatility σ^2 starting at 0, the simplest affine process, we get

$$\partial_{ss}\phi(u|s,x)|_{s=0} = 4^{-1}\sigma^4 u^4.$$

This means that the approximation errors of the estimates $\phi_n(u)$ and $\phi_{s,n}(u)$ based on local constant smoothing in x ($l = 0$), are of the order $(h_1^2 + h_2^1)\sigma^4 u^4/8$ and $(h_1 + h_1^{-1}h_2^1)\sigma^4 u^4/8$, respectively. So in order to be able to prove the uniform consistency in u we have to assume (20). In fact, the rate (22) can be proved to be optimal, provided the function $\psi(v|s,x)$ is at least two times differentiable in s and all partial derivatives in x up to order $l+1$ exist (see [19] for lower bounds for local polynomial estimates).

Let r_n be a sequence of positive r.v. and q_n be a sequence of positive real numbers. We shall write $r_n = O_{\text{a.s.}}(q_n)$ if there is a constant $D > 0$ such that $\mathrm{P}(\limsup_{n\to\infty} q_n^{-1} r_n \leq D) = 1$. In the case $\mathrm{P}(\limsup_{n\to\infty} q_n^{-1} r_n = 0) = 1$ we shall write $r_n = o_{\text{a.s.}}(q_n)$. Theorem 6.4 implies the following result on the strong uniform rates of convergence for the estimate $\rho_n^{(0)}$.

Theorem 6.6. *Suppose that the assumptions* (AX), (AN), (AG), (AP), (AK1) *and* (AK2) *hold. Let $\rho_n^{(0)}$ be the estimate for the transformed Lévy density $\rho^{(0)}$ defined in Section 4. If $\nu^{(0)} \in \mathfrak{L}_\varkappa$ for some $1 \leq \varkappa < 2$, then*

$$\|\rho^{(0)} - \rho_n^{(0)}\|_{L_\infty(\mathbb{R}^m)} = O_{\text{a.s.}}\left(\zeta_n \int_{[-U_n,U_n]^m} w^{-1}(\|v\|)\,\mathrm{d}v + U_n^{-(\varkappa-1)}\right),$$

with ζ_n being defined in (22).

6.3. Minimax upper bounds for $\widetilde{\rho}_n^{(0)}$

For any $1 \le \varkappa < 2$ let $\widetilde{\mathfrak{L}}_\varkappa$ stand for a class of Lévy densities $\nu^{(0)}$ satisfying

$$\left(\prod_{k=1}^{d} v_k\right)^{\varkappa} \mathcal{F}[\widetilde{\rho}^{(0)}](v) \le C, \quad v \in \mathbb{R}^m \tag{23}$$

for some $C > 0$, where the function $\widetilde{\rho}^{(0)}$ is related to the Lévy density $\widetilde{\nu}^{(0)}$ via (9). We need the following assumption concerning the asymptotic behavior of the sequence U_n.

(AH) The sequence U_n satisfies

$$U_n = O(n^{1-\delta}), \qquad M_n \sqrt{\frac{\log n}{n}} w^{-1}(U_n) = o(1), \quad n \to \infty$$

for some positive number δ, where

$$M_n = \sup_{\|v\| \le U_n} |\phi_\pi^{-1}(v)|.$$

Theorem 6.7. *Suppose that the assumptions* (AX), (AN), (AP), (AK1), (AK2) *and* (AH) *hold. Let $\widetilde{\rho}_n^{(0)}$ be the estimate for the transformed Lévy density $\widetilde{\rho}^{(0)}$ defined in Section 5 and let w be a monotone positive Lipschitz function on $\mathbb{R}_+$ satisfying*

$$0 < w(z) \le \log^{-1}(1+z), \quad z \in \mathbb{R}_+. \tag{24}$$

If $\widetilde{\nu}^{(0)} \in \widetilde{\mathfrak{L}}_\varkappa$ for some $1 \le \varkappa < 2$, then

$$\|\widetilde{\rho}^{(0)} - \widetilde{\rho}_n^{(0)}\|_{L_\infty(\mathbb{R}^m)} = O_{\text{a.s.}}\left(\sqrt{\frac{\log n}{n}} \int_{[-U_n,U_n]^m} w^{-1}(\|v\|)\mathfrak{R}(v)\,\mathrm{d}v + U_n^{-(\varkappa-1)}\right),$$

where $\mathfrak{R}(v) = |\phi_\pi(v)|^{-1} \int_{[-1,1]^m} |1 - \phi_\pi(v)/\phi_\pi(v+w)|\,\mathrm{d}w$.

Corollary 6.8. *Consider a class of affine models $\mathcal{A}(\bar{\sigma}, \varkappa)$, such that $\nu^{(0)} \in \widetilde{\mathfrak{L}}_\varkappa$ and*

$$|\phi_\pi(v)| \ge A \exp(-\bar{\sigma}^2 \|v\|_2^2), \quad v \in \mathbb{R}^m \tag{25}$$

with some constant $A > 0$. Then under the choice

$$U_n = m^{-1}\bar{\sigma}^{-1}\sqrt{\frac{1}{2}\log n + ((m+1)/2 + \varkappa - 1)\log\log n},$$

it holds

$$\|\widetilde{\rho}^{(0)} - \widetilde{\rho}_n^{(0)}\|_{L_\infty(\mathbb{R}^m)} = O_{\text{a.s.}}(\log^{-(\varkappa-1)/2} n).$$

Remark 6.9. Condition (25) holds if, for example, $\lambda_{\max}(\mathfrak{A}) \le \bar{\sigma}^2$ with $\mathfrak{A} := (\alpha_{ij}^{(0)})_{m+1 \le i,j \le d}$.

6.4. Lower risk bounds

The rates of Theorem 6.7 cannot be improved in general as the following theorem states

Proposition 6.10. *The following minimax lower risk bounds hold*

$$\liminf_{n\to\infty} \inf_{\widetilde{\rho}_n^{(0)}} \sup_{\chi\in\mathcal{A}(\bar{\sigma},\varkappa)} \mathrm{P}_\chi((\log n)^{(\varkappa-1)/2} \sup_{x\in\mathcal{D}} |\widetilde{\rho}^{(0)}(x) - \widetilde{\rho}_n^{(0)}(x)| > \varepsilon) > 0,$$

where ε is any positive number , $\widetilde{\rho}_n^{(0)}$ is any estimator of $\widetilde{\rho}^{(0)}$ based on n observations and the supremum is taken over all affine models χ from the class $\mathcal{A}(\bar{\sigma},\varkappa)$.

7. Numerical example

Let us consider a class of stochastic volatility models of the type

$$\begin{aligned} \mathrm{d}X(t) &= -\frac{1}{2}V(t)\,\mathrm{d}t + \sqrt{V(t)}\mathrm{d}W^S(t) + \mathrm{d}Z_t, \\ \mathrm{d}V(t) &= \lambda(\theta - V(t))\mathrm{d}t + \zeta\sqrt{V(t)}\mathrm{d}W^V(t), \end{aligned} \tag{26}$$

where W^S and W^V are two independent Brownian motions, λ, θ, ζ are positive constants and Z_t is a pure-jump Lévy process with Lévy density $\nu(x)$. This is a special type of the model introduced in [4] that satisfies our assumptions. In our numerical example we take $Z(t)$ to be an α-stable Lévy process with stability index $\alpha < 1$, i.e.,

$$\nu(x) = C/|x|^{1+\alpha}, \quad \rho(x) = 2\left(1 - \frac{\sin x}{x}\right)\nu(x)$$

for some constant $C > 0$. For the sake of simplicity we consider a fixed design and simulate a set of i.i.d. pairs

$$(X^{(k)}(0), X^{(k)}(\Delta)), \quad k = 1, \ldots, n, \tag{27}$$

with some fixed $\Delta > 0$, where

$$X^{(1)}(0) = \cdots = X^{(n)}(0) = 0.$$

Our aim is to reconstruct ρ using the sample (27). First, compute

$$\psi_{s,n}(v) := \frac{1}{n\Delta}\sum_{j=1}^{n}[\exp(\mathrm{i}vX^{(j)}(\Delta)) - \exp(\mathrm{i}vX^{(j)}(0))]$$

and

$$\Psi_{0,n}(v) := \int_{-1}^{1}(\psi_{s,n}(v) - \psi_{s,n}(v+w))\,\mathrm{d}w.$$

Remark 7.1. In the case of mixed-frequency data observations are usually available for different frequency scales Δ and the choice of an appropriate frequency for estimation procedure should be done depending on n, the number of points available for the given frequency scale. If Δ is too small then the variance of $\phi_{s,n}(v)$ explodes. On the other hand, if Δ is too large than the approximation error of $\phi_{s,n}(v)$ becomes large.

Next define a parametric family of functions

$$\rho_n^{(0)}(x;U) = \mathrm{Re}\left\{\frac{1}{2\pi}\int_{-U}^{U} \mathrm{e}^{-\mathrm{i}ux}[\Psi_{0,n}(u) - \Psi_{0,n}(U)]\,\mathrm{d}u\right\}, \quad U > 0$$

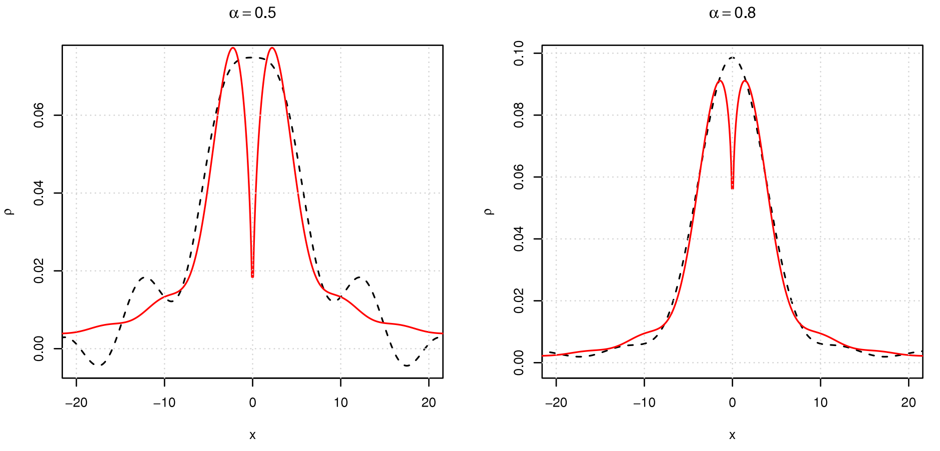

Fig. 1. Typical estimates for transformed Lévy density ρ (dashed black line) together with the true ρ (solid red line) in the Bates stochastic volatility model with symmetric stable process Z_t and different stability indexes α.

and find U by solving the following minimization problem

$$\widehat{U} = \operatorname{arginf}_U \left\{ \int_{\{|u|>U\}} |\Psi_{0,n}(u) - \Psi_{0,n}(U)|^2 \,\mathrm{d}u + \pi \int |\partial_{xx} \rho_n^{(0)}(x; U)| \,\mathrm{d}x \right\},$$

where $\pi > 0$ is a regularization parameter. In fact, this approach for choosing U employs additional information about the smoothness of ρ and turns out to be rather efficient in practice. In Fig. 1 typical results of estimation based on $n = 1000$ samples (27) with $\Delta = 0.1$ are shown for two specifications of the process Z_t. As can be seen the overall quality of estimation is good taking into account the severely ill-posedness of the underlying estimation problem. However, the behavior of the transformed Lévy density ρ at zero has not been captured by the estimation method. In order to correct $\rho(x)$ at $x = 0$, we separately estimate the stability index α using a modification of the spectral algorithm proposed in [6] for Lévy processes. Motivated by relations (2) and (3), we define for any $a \in (0, 1)$

$$\mathcal{O}(a) := \min_{(l_0,l_1,l_2,l_3)} \int_0^{\widehat{U}} (\psi_{s,n}(u|0) - l_3 u^a - l_2 u^2 - l_1 u - l_0)^2 \,\mathrm{d}u \tag{28}$$

and estimate α via $\widetilde{\alpha} := \operatorname{argmin}_{a\in(0,1)} \mathcal{O}(a)$. In Fig. 2 functions $\mathcal{O}(a)$ based on the same samples as in Fig. 1 are shown. The resulting estimates for α are $\widetilde{\alpha} = 0.451$ and $\widetilde{\alpha} = 0.783$ respectively. Now we correct the estimate $\rho_n^{(0)}(x; \widehat{U})$ by setting

$$\widehat{\rho}_n^{(0)}(x; \widehat{U}) = \begin{cases} c(\varepsilon)\left(1 - \dfrac{\sin x}{x}\right)|x|^{-(1+\widetilde{\alpha})}, & |x| \le \varepsilon, \\ \rho_n^{(0)}(x; \widehat{U}), & |x| > \varepsilon, \end{cases}$$

where for any $\varepsilon > 0$ the constant $c(\varepsilon)$ is chosen in such a way that function $\widehat{\rho}_n^{(0)}(x; \widehat{U})$ is continuous. Finally, we find a small enough $\varepsilon > 0$ which minimizes the integral $\int |\partial_{xx} \widehat{\rho}_n^{(0)}(x; \widehat{U})| \,\mathrm{d}x$. Here again the smoothness of ρ is used. A corrected estimate $\widehat{\rho}_n^{(0)}(x; \widehat{U})$ is shown in Fig. 3.

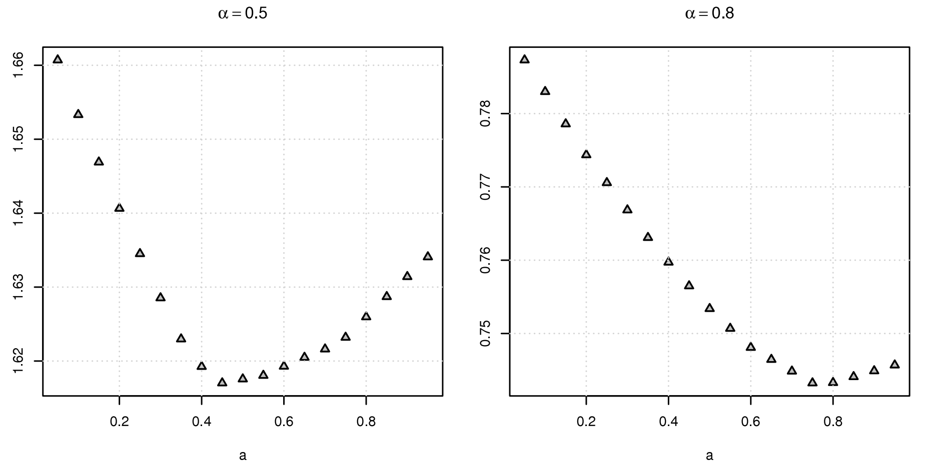

Fig. 2. Function $\mathcal{O}(a)$ in the case of symmetric stable process Z_t with stability indexes 0.5 (left) and 0.8 (right) respectively.

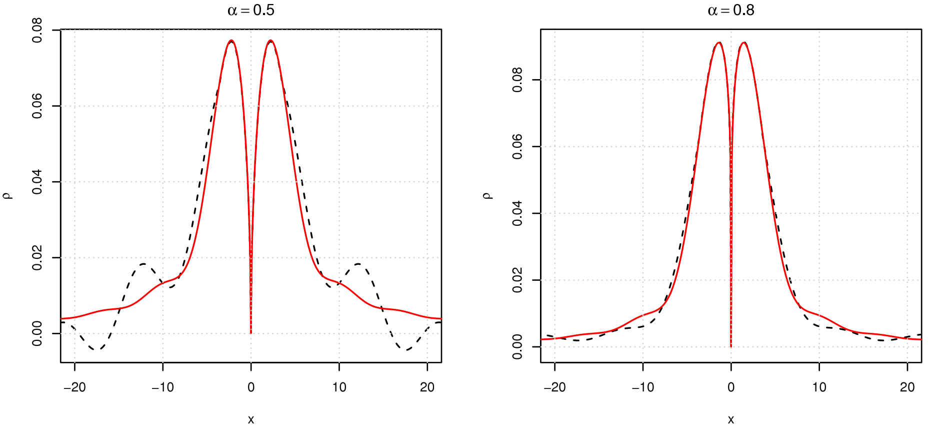

Fig. 3. Corrected estimates for transformed Lévy density ρ (dashed black line) together with the true ρ (solid red line) in the Bates stochastic volatility model with symmetric stable process Z_t and different stability indexes α.

8. Exponential inequalities for dependent sequences

The following theorem can be found in [15].

Theorem 8.1. *Let $(Z_k,\ k \geq 1)$ be a strongly mixing sequence of centered real-valued random variables on the probability space $(\Omega, \mathcal{F}, P)$ with the mixing coefficients satisfying*

$$\alpha(n) \leq \bar{\alpha} \exp(-cn), \quad n \geq 1,\ \bar{\alpha} > 0, \quad c > 0. \tag{29}$$

Assume that $\sup_{k \geq 1} |Z_k| \leq M$ a.s., then there is a positive constant C depending on c and $\bar{\alpha}$ such that

$$\mathrm{P}\left\{\sum_{i=1}^{n} Z_i \geq \zeta\right\} \leq \exp\left[-\frac{C\zeta^2}{nv^2 + M^2 + M\zeta \log^2(n)}\right]$$

for all $\zeta > 0$ *and* $n \geq 4$, *where*

$$v^2 = \sup_i \left(\mathrm{E}[Z_i]^2 + 2 \sum_{j \geq i} \mathrm{Cov}(Z_i, Z_j) \right).$$

Corollary 8.2. *Denote*

$$\rho_j = \mathrm{E}[Z_j^2 \log^{2(1+\varepsilon)}(|Z_j|^2)], \quad j = 1, 2, \ldots,$$

with arbitrary small $\varepsilon > 0$ *and suppose that all* ρ_j *are finite. Then*

$$\sum_{j \geq i} \mathrm{Cov}(Z_i, Z_j) \leq C \max_j \rho_j$$

for some constant $C > 0$, *provided* (29) *holds. Consequently the following inequality holds*

$$v^2 \leq \sup_i \mathrm{E}[Z_i]^2 + C \max_j \rho_j.$$

Proof. Due to the Rio inequality

$$|\mathrm{Cov}(Z_i, Z_j)| \leq 2 \int_0^{\alpha(|j-i|)} Q_{Z_i}(u) Q_{Z_j}(u) \mathrm{d}u$$

where for any random variable X we denote by Q_X the quantile function of X. Define

$$\rho_X = \mathrm{E}[X^2 \log^{2(1+\varepsilon)}(|X|^2)].$$

The Markov inequality implies for small enough $u > 0$

$$\begin{aligned} \mathrm{P}\left(|X| > \frac{\rho_X^{1/2}}{u^{1/2} |\log(u)|^{(1+\varepsilon)}} \right) &\leq \mathrm{E}\left[X^2 \log^{2(1+\varepsilon)}(|X|^2) \right] \frac{\rho_X^{-1}}{u^{-1} \log^{-2(1+\varepsilon)}(u)} \\ &\quad \times \log^{-2(1+\varepsilon)} \left(\frac{\rho_X}{u \log^{2(1+\varepsilon)}(u)} \right) \\ &= u \log^{-2(1+\varepsilon)}(\rho_X \log^{-2(1+\varepsilon)}(u)) \leq u \end{aligned}$$

and therefore

$$Q_X(u) \leq \frac{\rho_X^{1/2}}{u^{1/2} |\log(u)|^{(1+\varepsilon)}}.$$

Hence

$$|\mathrm{Cov}(Z_i, Z_j)| \leq 2 \int_0^{\alpha(|j-i|)} \frac{\sqrt{\rho_i \rho_j}}{u \log^{2(1+\varepsilon)}(u)} \mathrm{d}u \leq 2 \sqrt{\rho_i \rho_j} \log^{-1-2\varepsilon}(\alpha(|j-i|))$$

and

$$\sum_{j \geq i} \mathrm{Cov}(Z_i, Z_j) \leq C \sqrt{\rho_i \rho_j} \sum_{j > i} \frac{1}{|j-i|^{1+2\varepsilon}}$$

with some constant $C > 0$ depending on $\bar{\alpha}$. $\square$

8.1. Bounds on large deviation probabilities for weighted sup norms

Let Z_j, $j = 1, \dots, n$, be a sequence of random vectors in $\mathbb{R}^d$ and let $G_n(u, z)$, $n = 1, 2, \dots$, be a sequence of complex-valued functions defined on $\mathbb{R}^d \times \mathbb{R}^d$ for some natural number d. Define

$$m_n(u) = \frac{1}{n} \sum_{j=1}^{n} G_n(u, Z_j).$$

Proposition 8.3. *Suppose that the following assumptions hold:*

(AZ1) *The sequence Z_j, $j = 1, \dots, n$, is strictly stationary and is α-mixing with the mixing coefficients $(\alpha_Z(k))_{k \in \mathbb{N}}$ satisfying*

$$\alpha_Z(k) \le \bar{\alpha}_0 \exp(-\bar{\alpha}_1 k), \quad k \in \mathbb{N},$$

for some $\bar{\alpha}_0 > 0$ and $\bar{\alpha}_1 > 0$.

(AZ2) *It holds $\mathrm{E}\,\|Z_j\|^p < \infty$ for some $p > 2$.*

(AG1) *Each function $G_n(u, z)$, $n \in \mathbb{N}$, is Lipschitz in u with (at most) linearly growing (in z) Lipschitz constant, i.e., for any $u_1, u_2 \in \mathbb{R}^d$*

$$|G_n(u_1, z) - G_n(u_2, z)| \le L_n(a + b\|z\|)\|u_1 - u_2\|$$

where a, b are two non-negative real numbers not depending on n and the sequence L_n does not depend on u.

(AG2) *There are two sequences μ_n and σ_n, such that*

$$\mathrm{E}[|G_n(u, Z)|^2] \le \sigma_n^2, \qquad |G_n(u, z)| \le \mu_n, \quad (u, z) \in \mathbb{R}^{2d}$$

and the following asymptotic relations are fulfilled

$$\mu_n/\sigma_n^2 = O(1), \qquad \mu_n/\sigma_n = O(n^{1/2}\log^{-4} n),$$
$$\sigma_n^2 = O(n), \qquad L_n = O(\mu_n),$$

as $n \to \infty$.

Let w be a Lipschitz continuous, positive, monotone decreasing on the $\mathbb{R}_+$ function such that

$$0 < w(z) \le \log^{-1/2}(e + z), \quad z \in \mathbb{R}_+. \tag{30}$$

Then there is $\delta' > 0$ and $\xi_0 > 0$, such that the inequality

$$\mathrm{P}\left\{ \log^{-(1+\varepsilon)}(1 + \mu_n) \sqrt{\frac{n}{\sigma_n^2 \log n}} \sup_{u \in \mathbb{R}^d} [w(\|u\|)|m_n(u) - \mathrm{E}[m_n(u)]|] > \xi \right\}$$
$$\le B n^{-1-\delta'} \tag{31}$$

holds for any $\xi > \xi_0$, some positive constant B depending on ξ and arbitrary small $\varepsilon > 0$.

Proof. Define for $j = 1, \dots, n$,

$$\zeta_j(u) = G_n(u, Z_j) - \mathrm{E}[G_n(u, Z_j)]$$

and introduce the process

$$\mathcal{W}_n(u) = \frac{1}{n} \sum_{j=1}^{n} w(\|u\|)\zeta_j(u).$$

Consider the sequence $A_k = e^k$, $k \in \mathbb{N}$ and cover each cube $[-A_k, A_k]^d$ with $M_k = (\lfloor (2d^{1/2}A_k)/\gamma \rfloor + 1)^d$ disjoint small cubes $\Lambda_{k,1}, \ldots, \Lambda_{k,M_k}$, the length of each cube being equal to $\gamma/d^{1/2}$. Let $u_{k,1}, \ldots, u_{k,M_k}$ be the centers of these cubes. We have for any natural $K > 0$

$$\max_{k=1,\ldots,K} \sup_{A_{k-1}<\|u\|\le A_k} |\mathcal{W}_n(u)| \le \max_{k=1,\ldots,K} \max_{\|u_{k,m}\|>A_{k-1}} |\mathcal{W}_n(u_{k,m})| + \max_{k=1,\ldots,K} \max_{1\le m\le M_k} \sup_{u\in\Lambda_{k,m}} |\mathcal{W}_n(u) - \mathcal{W}_n(u_{k,m})|.$$

Hence

$$\mathrm{P}(\max_{k=1,\ldots,K} \sup_{A_{k-1}<\|u\|\le A_k} |\mathcal{W}_n(u)| > \lambda) \le \sum_{k=1}^{K} \sum_{\{\|u_{k,m}\|>A_{k-1}\}} \mathrm{P}(|\mathcal{W}_n(u_{k,m})| > \lambda/2) + \mathrm{P}(\sup_{\|u-v\|<\gamma} |\mathcal{W}_n(v) - \mathcal{W}_n(u)| > \lambda/2). \tag{32}$$

It holds for any $u, v \in \mathbb{R}^d$

$$\begin{aligned} |\mathcal{W}_n(v) - \mathcal{W}_n(u)| &\le 2\mu_n |w(\|v\|) - w(\|u\|)| + \frac{1}{n}\sum_{j=1}^{n} |G_n(v, Z_j) - G_n(u, Z_j)| \\ &\quad + \frac{1}{n}\sum_{j=1}^{n} |\mathrm{E}[G_n(v, Z_j) - G_n(u, Z_j)]| \\ &\le 2(L_n \vee \mu_n)\|u - v\| \left[L_w + \frac{1}{n}\sum_{j=1}^{n}(a + b\|Z_j\|) + \frac{1}{n}\sum_{j=1}^{n}(a + b\,\mathrm{E}\,\|Z_j\|) \right], \end{aligned} \tag{33}$$

where L_ω is the Lipschitz constant of w. By the Markov inequality

$$\mathrm{P}\left(\frac{1}{n}\sum_{j=1}^{n}[\|Z_j\| - \mathrm{E}\,\|Z_j\|] > c\right) \le c^{-p} n^{-p}\,\mathrm{E}\left|\sum_{j=1}^{n}[\|Z_j\| - \mathrm{E}\,\|Z_j\|]\right|^p$$

for any $c > 0$. Using the moment inequality of Yokoyama (see [23]), we get

$$\mathrm{E}\left|\sum_{j=1}^{n}[\|Z_j\| - \mathrm{E}\,\|Z_j\|]\right|^p \le C_p(\alpha) n^{p/2},$$

where $C_p(\alpha)$ is some constant depending on p and $\alpha = (\bar{\alpha}_0, \bar{\alpha}_1)$ from the assumption (AZ1). Hence

$$\mathrm{P}\left(\frac{1}{n}\sum_{j=1}^{n}\|Z_j\| > 2\beta_1\right) \le C_p(\alpha) n^{-p/2}/\beta_1^p \tag{34}$$

with $\beta_1 = \mathrm{E}\,\|Z_j\|$. Setting

$$\gamma = \lambda/(4(L_n \vee \mu_n)(2a + 3b\beta_1 + L_\omega))$$

and combining (33) with the inequality (34), we obtain

$$\mathrm{P}(\sup_{\|u-v\|<\gamma} |\mathcal{W}_n(v) - \mathcal{W}_n(u)| > \lambda/2) \le B_1 n^{-p/2} \tag{35}$$

with some constant B_1 depending neither on λ or n. Turn now to the first term on the right-hand side of (32). If $\|u_{k,m}\| > A_{k-1}$, then it follows from Theorem 8.1 and Corollary 8.2

$$
\begin{aligned}
&\mathrm{P}(|\operatorname{Re}[\mathcal{W}_n(u_{k,m})]| > \lambda/4) \\
&\quad \le B_2 \exp\left(-\frac{\lambda^2 n}{B_3 w^2(A_{k-1})\sigma_n^2 \log^{2(1+\varepsilon)}(1+\mu_n) + w^2(A_{k-1})\mu_n^2/n + \lambda\mu_n \log^2(n) w(A_{k-1})}\right), \\
&\mathrm{P}(|\operatorname{Im}[\mathcal{W}_n(u_{k,m})]| > \lambda/4) \\
&\quad \le B_4 \exp\left(-\frac{\lambda^2 n}{B_3 w^2(A_{k-1})\sigma_n^2 \log^{2(1+\varepsilon)}(1+\mu_n) + w^2(A_{k-1})\mu_n^2/n + \lambda\mu_n \log^2(n) w(A_{k-1})}\right)
\end{aligned}
$$

with some constants B_2 , B_3 and B_4 depending only on the characteristics of the process Z, and $\varepsilon > 0$. Indeed, due to (AG2) it holds

$$
\begin{aligned}
\mathrm{E}[|\zeta_j|^2 \log^{2(1+\varepsilon)}(|\zeta_j|^2)\mathbf{1}_{\{|\zeta_j|>1\}}] &\le \mathrm{E}[|G_n(u, Z_j)|^2] \log^{2(1+\varepsilon)}(2\mu_n) \\
&\le B_5 \sigma_n^2 \log^{2(1+\varepsilon)}(1+\mu_n)
\end{aligned}
$$

with some constant $B_5 > 0$. Hence

$$
\begin{aligned}
\mathrm{E}[|w\zeta_j|^2 \log^{2(1+\varepsilon)}(|w\zeta_j|^2)] &= w^2\, \mathrm{E}[|\zeta_j|^2 \log^{2(1+\varepsilon)}(|\zeta_j|^2)] + w^2 \log^{2(1+\varepsilon)}(w^2)\, \mathrm{E}[|\zeta_j|^2] \\
&\le w^2(1 + B_5\sigma_n^2 \log^{2(1+\varepsilon)}(1+\mu_n)) + w^2 \log^{2(1+\varepsilon)}(w^2)\sigma_n^2.
\end{aligned}
$$

Taking $\lambda = \xi n^{-1/2}\sigma_n \log^{1/2}(n) \log^{1+\varepsilon}(1+\mu_n)$ with some $\xi > 0$, we get

$$
\begin{aligned}
&\sum_{\{\|u_{k,m}\|>A_{k-1}\}} \mathrm{P}(|\mathcal{W}_n(u_{k,m})| > \lambda/2) \le (\lfloor (2d^{1/2}A_k)/\gamma \rfloor + 1)^d \\
&\quad \times \exp\left(-\frac{\lambda^2 n}{B_3 w^2(A_{k-1})\sigma_n^2 \log^{2(1+\varepsilon)}(1+\mu_n) + w^2(A_{k-1})\mu_n^2/n + \lambda\mu_n \log^2(n) w(A_{k-1})}\right) \\
&\quad \lesssim A_k^d n^{d/2} \sigma_n^{-1}(L_n \vee \mu_n) \log^{-3/2-\varepsilon}(n) \exp\left(-\frac{B\xi^2 \log(n)}{w^2(A_{k-1})}\right), \quad n \to \infty
\end{aligned}
$$

with some constant $B > 0$. Fix $\theta > 0$ such that $B\theta > 1$ and compute

$$
\begin{aligned}
&\sum_{\{\|u_{k,m}\|>A_{k-1}\}} \mathrm{P}(|\mathcal{W}_n(u_{k,m})| > \lambda/2) \\
&\quad \lesssim \mathrm{e}^{\mathrm{d}k - \theta B(k-1)} n^{(d+1)/2} \log^{-3/2-\varepsilon}(n) \mathrm{e}^{-B(k-1)(\xi^2 \log n - \theta)} \\
&\quad \lesssim \mathrm{e}^{k(1-\theta B)} \log^{-3/2-\varepsilon}(n) \mathrm{e}^{-B(k-1)(\xi^2 \log n - \theta) + (d+1)\log(n)/2}.
\end{aligned}
$$

If $\xi^2 \log n > \theta$ we derive

$$
\sum_{k=2}^{K} \sum_{\{\|u_{k,m}\|>A_{k-1}\}} \mathrm{P}(|\mathcal{W}_n(u_{k,m})| > \lambda/2) \lesssim \log^{-3/2-\varepsilon}(n) \mathrm{e}^{-(B\xi^2 - (d+1)/2)\log(n)}.
$$

Taking $\xi > \xi_0$ for large enough ξ_0, we get (31). $\square$

9. Proofs

9.1. Proof of Theorem 6.4

The smallest eigenvalue λ_Γ of the matrix Γ satisfies

$$\begin{aligned}\lambda_\Gamma &= \min_{\|W\|=1} W^\top \Gamma W \\ &\geq \min_{\|W\|=1} W^\top \mathrm{E}[\Gamma]W + \min_{\|W\|=1} W^\top(\Gamma - \mathrm{E}[\Gamma])W \\ &\geq \min_{\|W\|=1} W^\top \mathrm{E}[\Gamma]W - \sum_{k=0}^{2}\sum_{|\mathbf{m}_1|,|\mathbf{m}_2|\leq l} |\Gamma_{k,\mathbf{m}_1,\mathbf{m}_2} - \mathrm{E}[\Gamma_{k,\mathbf{m}_1,\mathbf{m}_2}]|. \end{aligned}\tag{36}$$

By the assumption (AG)

$$\min_{\|W\|=1} [W^\top \mathrm{E}[\Gamma]W] \geq \gamma_0$$

with some $\gamma_0 > 0$. For $j = 1, \dots, n$, any $k = 0, 1, 2$ and any multi-indices $\mathbf{m}_1, \mathbf{m}_2$ such that $|\mathbf{m}_1|, |\mathbf{m}_2| \leq l$, define

$$\begin{aligned}\Delta_j &= \frac{1}{h_1 h_2^m}\left(\frac{\delta_j}{h_1}\right)^k \left(\frac{X^{(m)}(t_j) - x}{h_2}\right)^{\mathbf{m}_1+\mathbf{m}_2} K\left(\frac{\delta_j}{h_1}, \frac{X^{(m)}(t_j) - x}{h_2}\right) \\ &\quad - \int_{\mathbb{R}^m}\int_{\mathbb{R}_+} t^k z^{\mathbf{m}_1+\mathbf{m}_2} K(t,z) p_{m,\pi}(x + h_2 z) p_\delta(h_1 t)\,\mathrm{d}t\,\mathrm{d}z.\end{aligned}$$

We have $\mathrm{E}[\Delta_j] = 0$,

$$|\Delta_j| \leq h_1 h_2^{-m} \sup_{t\in\mathbb{R}_+}\sup_{z\in\mathbb{R}^m} [(1+t^2)(1+\|z\|^{2l})K(t,z)] =: K_1 h_1 h_2^{-m}$$

and

$$\begin{aligned}\mathrm{E}[\Delta_j]^2 &\leq \int_{\mathbb{R}^m}\int_{\mathbb{R}_+} t^{2k} z^{2u_1+2u_2} K^2(t,z) p_{m,\pi}(x+h_2 z) p_\delta(h_1 t)\,\mathrm{d}t\,\mathrm{d}z \\ &\leq \frac{\bar{p}_{m,\pi}\bar{p}_\delta}{h_1 h_2^m}\int_{\mathbb{R}^m}(1+t^4)(1+\|z\|^{4l})K^2(t,z)\,\mathrm{d}z =: K_2 h_1^{-1} h_2^{-m},\end{aligned}$$

where $\bar{p}_X = \sup_{z\in\mathbb{R}^m} p_{m,\pi}(z)$, $\bar{p}_\delta = \sup_{s\in[0,T]} p_\delta(s)$ and K_1, K_2 are two positive constants. According to Proposition 8.3, we have for any $\zeta > 0$

$$\mathrm{P}(|\Gamma_{k,\mathbf{m}_1,\mathbf{m}_2} - \mathrm{E}[\Gamma_{k,\mathbf{m}_1,\mathbf{m}_2}]| \geq \zeta) = \mathrm{P}\left(\frac{1}{n}\left|\sum_{j=1}^n \Delta_j\right| \geq \zeta\right) \leq D_0 \exp(-\zeta B_0 n h_1 h_2^m) \tag{37}$$

with some positive constants D_0 and B_0. Combining (36) with (37), we get

$$\mathrm{P}(\lambda_\Gamma \leq \gamma_0/2) \leq D_0 N_l^2 \exp(-\gamma_0 B_0 n h_1 h_2^m / 2N_l^2),$$

where N_l^2 is the number of elements in the matrix Γ. Introduce the matrices $Q_k = (Q_{k,j,\mathbf{m}})_{1\leq j\leq n,\,|\mathbf{m}|\leq l}$ with elements

$$Q_{k,j,\mathbf{m}} = \left(\frac{\delta_j}{h_1}\right)^k \left(\frac{X^{(m)}(t_j) - x}{h_2}\right)^{\mathbf{m}} \sqrt{\frac{1}{n h_1 h_2^m} K\left(\frac{\delta_j}{h_1}, \frac{X^{(m)}(t_j) - x}{h_2}\right)}.$$

Set

$$\phi_m(v|s,x) = \phi(v,0\ldots,0|s,\mathrm{E}[X_{m+1}(0)],\ldots,\mathrm{E}[X_d(0)]).$$

Denote by $Q_{k,\mathbf{m}}$ the $\mathbf{m}$th column of Q_k, $k=0,1$, and define

$$Q_0^C = \sum_{|\mathbf{m}|\le l} \frac{\partial_{x^{\mathbf{m}}}\phi_m(v|0,x)h_2^{\mathbf{m}}}{\mathbf{m}!} Q_{0,\mathbf{m}},$$

$$Q_1^C = h_1 \sum_{|\mathbf{m}|\le l} \frac{\partial_{x^{\mathbf{m}}}\partial_s\phi_m(v|0,x)h_2^{\mathbf{m}}}{\mathbf{m}!} Q_{1,\mathbf{m}},$$

$$Q = \begin{pmatrix} Q_0 \\ Q_1 \end{pmatrix}, \qquad Q^C = \begin{pmatrix} Q_0^C \\ Q_1^C \end{pmatrix}.$$

Since $\Gamma = Q^\top Q$, we get $(P(0),P(0))^\top \Gamma^{-1} Q^\top Q^C = (\phi_m(v|0,x), h_1\phi_m(v|0,x))$. Thus, we can write

$$\begin{aligned}(\phi_n(v)-\phi_m(v|0,x), h_1(\phi_{s,n}(v)-\phi_{m,s}(v|0,x)))^\top &= (P(0),P(0))^\top\Gamma^{-1}(S-Q^\top Q^C)\\ &= (P(0),P(0))^\top\Gamma^{-1}\varepsilon_n(v),\end{aligned}$$

where $\varepsilon_n(v) = (\varepsilon_{n,0}(v),\varepsilon_{n,1}(v))$ and

$$\begin{aligned}\varepsilon_{n,k,\mathbf{m}}(v) = \frac{1}{nh_1h_2^m}\sum_{j=1}^n &[Z_j(v) - \bar{\mathcal{Q}}_0(X^{(m)}(t_j)-x) - \delta_j\bar{\mathcal{Q}}_1(X^{(m)}(t_j)-x)]\\ &\times \left(\frac{\delta_j}{h_1}\right)^k \left(\frac{X^{(m)}(t_j)-x}{h_2}\right)^{\mathbf{m}} K\left(\frac{\delta_j}{h_1}, \frac{X^{(m)}(t_j)-x}{h_2}\right),\end{aligned}$$

where

$$\bar{\mathcal{Q}}_0(z) = \sum_{|\mathbf{m}|\le l}\frac{\partial_{x^{\mathbf{m}}}\phi_m(v|0,x)z^{\mathbf{m}}}{\mathbf{m}!}$$

and

$$\bar{\mathcal{Q}}_1(z) = \sum_{|\mathbf{m}|\le l}\frac{\partial_{x^{\mathbf{m}}}\partial_s\phi_m(v|0,x)z^{\mathbf{m}}}{\mathbf{m}!}.$$

So, on the set $\{\lambda_\Gamma > \gamma_0/2\}$ we get

$$\begin{aligned}\max\{|\phi_n(v)-\phi_m(v|0,x)|, |\phi_{s,n}(v)-\phi_{m,s}(v|0,x)|\} &\le \|\Gamma\varepsilon_n\| \le \lambda_\Gamma^{-1}\|\varepsilon_n\|\\ &\le 2\gamma_0^{-1}\|\varepsilon_n\|\\ &\le 2\gamma_0^{-1}N_l^{1/2}\max_{k,\mathbf{m}}|\varepsilon_{n,k,\mathbf{m}}(v)|.\end{aligned}$$

Denote

$$\begin{aligned}\Delta_{j,k,\mathbf{m}}^{(1)}(v) := h_1^{-1}h_2^{-m}&[Z_j(v) - \phi_m(v|\delta_j, X^{(m)}(t_j))]\\ &\times\left(\frac{\delta_j}{h_1}\right)^k \left(\frac{X^{(m)}(t_j)-x}{h_2}\right)^{\mathbf{m}} K\left(\frac{\delta_j}{h_1}, \frac{X^{(m)}(t_j)-x}{h_2}\right),\end{aligned}$$

$$\Delta^{(2)}_{j,k,\mathbf{m}}(v) := h_1^{-1}h_2^{-m}[\phi_m(v|\delta_j, X^{(m)}(t_j)) - \bar{\mathcal{Q}}_0(X^{(m)}(t_j)-x) - \delta_j\bar{\mathcal{Q}}_1(X^{(m)}(t_j)-x)] \times \left(\frac{\delta_j}{h_1}\right)^k \left(\frac{X^{(m)}(t_j)-x}{h_2}\right)^{\mathbf{m}} K\left(\frac{\delta_j}{h_1}, \frac{X^{(m)}(t_j)-x}{h_2}\right).$$

It holds

$$|\varepsilon_{n,k,\mathbf{m}}(v)| \le \left|\frac{1}{n}\sum_{j=1}^n \Delta^{(1)}_{j,k,\mathbf{m}}(v)\right| + \left|\frac{1}{n}\sum_{j=1}^n [\Delta^{(2)}_{j,k,\mathbf{m}}(v) - \mathrm{E}\,\Delta^{(2)}_{j,k,\mathbf{m}}(v)]\right| + |\mathrm{E}\,\Delta^{(2)}_{j,k,\mathbf{m}}(v)|.$$

Note that $\mathrm{E}[\Delta^{(1)}_{j,k,\mathbf{m}}(v)] = 0$ and

$$|\Delta^{(1)}_{j,k,\mathbf{m}}(v)| \le A_{11}h_1^{-1}h_2^{-m}, \qquad \mathrm{E}[\Delta^{(1)}_{j,k,\mathbf{m}}(v)]^2 \le A_{12}h_1^{-1}h_2^{-m},$$
$$|\Delta^{(2)}_{j,k,\mathbf{m}}(v) - \mathrm{E}[\Delta^{(2)}_{j,k,\mathbf{m}}(v)]| \le A_{21}h_2^{l-m}, \qquad \mathrm{E}[\Delta^{(2)}_{j,k,\mathbf{m}}(v)]^2 \le A_{22}h_2^{2l-m}$$

with some positive constants A_{11} , A_{12} , A_{21} and A_{22} not depending on x. Proposition 8.3 implies that

$$\mathrm{P}\left(\sup_{v\in\mathbb{R}^m}\left\{w(\|v\|)\left|\frac{1}{n}\sum_{j=1}^n \Delta^{(1)}_{j,k,\mathbf{m}}(v)\right|\right\} > A_1\frac{\log^{3/2+\varepsilon} n}{\sqrt{nh_1h_2^m}}\right) \le B_1 n^{-1-\delta'}$$

and

$$\mathrm{P}\left(\sup_{v\in\mathbb{R}^m}\left\{w(\|v\|)\left|\frac{1}{n}\sum_{j=1}^n (\Delta^{(2)}_{j,k,\mathbf{m}}(v) - \mathrm{E}[\Delta^{(2)}_{j,k,\mathbf{m}}(v)])\right|\right\} > A_2\frac{\log^{3/2+\varepsilon} n}{\sqrt{nh_2^{2l-m}}}\right) \le B_2 n^{-1-\delta'}$$

for some positive constants A_1, B_1, A_2 and B_2 not depending on k and $\mathbf{m}$. Due to Lemma A.1 and Assumption (AN), the function $\phi_m(v|s,x)$ is at least two times differentiable in s. Hence

$$\sup_{|t|\le h_1}\sup_{\|z-x\|\le h_2} |\phi_m(v|0,z) - \bar{\mathcal{Q}}_0(z-x) - t\bar{\mathcal{Q}}_1(z-x)|$$
$$\le C\Bigg[h_1^2 \sup_{|t|\le h_1}\sup_{\|z-x\|\le h_2} |\partial_{ss}\phi_m(v|t,x)| + h_2^{l+1}\max_{|\mathbf{m}|\le l+1}\sup_{\|z-x\|\le h_2}|\partial_{x^{\mathbf{m}}}\phi_m(v|0,x)|$$
$$+ h_1h_2^{l+1}\max_{|\mathbf{m}|\le l+1}\sup_{\|z-x\|\le h_2}|\partial_{x^{\mathbf{m}}}\partial_s\phi_m(v|0,x)|\Bigg]$$

for some constant $C > 0$. Using Lemma A.1, we get $|\mathrm{E}\,\Delta^{(2)}_{j,k,\mathbf{m}}(v)| \le A_3(h_1^2 + h_2^{l+1}) \|v\|_2^{2(l+2)}$, $\|v\|_2 > 1$.

9.2. *Proof of Theorem 6.6*

Theorem 6.4 implies

$$\mathrm{P}(\sup_{v\in\mathbb{R}^m}[w(\|v\|)|\Psi_0(v) - \Psi_{0,n}(v)|] > 2^{m+1}A\zeta_n) \le Bn^{-1-\delta}. \tag{38}$$

Using the identity $\mathcal{L} = \Psi_0(v) - \mathcal{F}[\rho^{(0)}](v)$, we get

$$\mathcal{L}_n - \mathcal{L} = \int_{\mathbb{R}^m} U_n^{-m}\mathcal{K}_{\mathcal{L}}(v/U_n)(\Psi_{0,n}(v) - \Psi_0(v))\,\mathrm{d}v$$

$$+ \int_{\mathbb{R}^m} U_n^{-m} \mathcal{K}_{\mathcal{L}}(v/U_n) \mathcal{F}[\rho^{(0)}](v)\, \mathrm{d}v$$

and hence

$$|\mathcal{L}_n - \mathcal{L}| \le \int_{\mathbb{R}^m} U_n^{-m} |\mathcal{K}_{\mathcal{L}}(v/U_n)| |\Psi_{0,n}(v) - \Psi_0(v)|\, \mathrm{d}u$$
$$+ U_n^{-m} \int_{\mathbb{R}^m} |\mathcal{K}_{\mathcal{L}}(v/U_n) \mathcal{F}[\rho^{(0)}](v)|\, \mathrm{d}v.$$

By the inequality (23)

$$\begin{aligned} \int_{\mathbb{R}^m} |\mathcal{K}_{\mathcal{L}}(v/U_n) \mathcal{F}[\rho^{(0)}](v)|\, \mathrm{d}v &\le C \int_{\mathbb{R}^m} [|v_1|^{-\varkappa} \times \cdots \times |v_d|^{-\varkappa}] |\mathcal{K}_{\mathcal{L}}(v/U_n)|\, \mathrm{d}v \\ &\le C_1 U_n^{-m(\varkappa - 1)} \end{aligned} \tag{39}$$

with some constant $C_1 > 0$. Combining (39) with (38), we get

$$U_n^m |\mathcal{L}_n - \mathcal{L}| = O_{a.s} \left(\zeta_n \int_{[-U_n, U_n]^m} w^{-1}(\|v\|)\, \mathrm{d}v + U_n^{-m(\varkappa - 1)} \right).$$

Finally, using the Fourier inversion formula, we derive

$$\begin{aligned} \sup_{z \in \mathbb{R}^m} |\rho^{(0)}(z) - \rho_n^{(0)}(z)| &\le \left[\int_{\mathbb{R}^m} |\mathcal{K}_\rho(v/U_n)| |\Psi_{0,n}(v) - \Psi_0(v)|\, \mathrm{d}v + U_n^m |\mathcal{L}_n - \mathcal{L}| \right] \\ &\quad + \left| \int_{\{\|v\| > a_K U_n\}} \mathcal{F}[\rho^{(0)}](v) (1 - \mathcal{K}_\rho(v/U_n))\, \mathrm{d}v \right|. \end{aligned}$$

9.3. *Proof of Theorem 6.7*

Fix some $D > 0$ and consider the event

$$\mathcal{A} = \left\{ \sup_{\|v\| \le U_n} [w(\|v\|) |\phi(v) - \phi_n(v)|] \le D \sqrt{\frac{\log n}{n}} \right\}.$$

By the assumption (AH), it holds on $\mathcal{A}$

$$\sup_{\|v\| < U_n} \left| \frac{\phi(v) - \phi_n(v)}{\phi(v)} \right| \le D M_n w^{-1}(U_n) \sqrt{\frac{\log n}{n}} = o(1), \quad n \to \infty$$

and hence

$$\log \phi_n(v) - \log \phi(v) = \log \left[1 + \frac{\phi_n(v) - \phi(v)}{\phi(v)} \right] = \frac{\phi_n(v) - \phi(v)}{\phi(v)} + \mathcal{R}(v)$$

with

$$|\mathcal{R}(v)| \le c \left| \frac{\phi_n(v) - \phi(v)}{\phi(v)} \right|^2$$

for some constant $c > 0$ and all v satisfying $\|v\| \le U_n$. On the other hand, Proposition 8.3 (take $G(u, z) = \exp(\mathrm{i} u^\top z)$) implies

$$\mathrm{P}(\bar{\mathcal{A}}) \le B' n^{-1-\delta'}, \quad n \to \infty$$

for some $B' > 0$ and $\delta' > 0$, provided D is large enough. Therefore it holds on $\mathcal{A}$

$$w(\|v\|)|\Psi_{\pi,n}(v) - \Psi_\pi(v)t| \le C\sqrt{\frac{\log n}{n}}|\phi_\pi(v)|^{-1}\int_{[-1,1]^m}|1-\phi_\pi(v)/\phi_\pi(v+w)|\,\mathrm{d}w$$

for some constant $C > 0$ and all v satisfying $\|v\| \le U_n$. The rest of the proof is similar to the proof of Theorem 6.6.

9.4. *Proof of Proposition 6.10*

In order to prove minimax lower bounds we apply general results from [21]. Let Θ be a semi-parametric class of models. Consider a family $\{P_\theta, \theta \in \Theta\}$ of probability measures, indexed by Θ. For any $\theta_1, \theta_2 \in \Theta$ let $d(\theta_1, \theta_2)$ be a semi-distance between two models θ_1 and θ_2.

Lemma 9.1. *Suppose that Θ contains two elements θ_1 and θ_2 such that $d(\theta_1,\theta_2) > 2s$ for some $s > 0$ and $\chi^2(P_{\theta_1}^{\otimes n}, P_{\theta_2}^{\otimes n}) \le \tau < 1/2$, where*

$$\chi^2(P,Q) =: \begin{cases} \int\left(\frac{\mathrm{d}P}{\mathrm{d}Q}-1\right)^2\mathrm{d}Q, & \text{if } P \ll Q \\ +\infty, & \text{otherwise}\end{cases}$$

for any two measures P and Q. Then

$$\inf_{\widehat{\theta}}\sup_{\theta\in\Theta} \mathrm{P}_\theta(d(\widehat{\theta},\theta) \ge s) \ge c(\tau) > 0,$$

where $c(\tau)$ is a constant depending on τ and the infimum is taken over all estimates $\widehat{\theta}$ of θ based on N observations under P_θ.

Turn now to the construction of models θ_1 and θ_2 from the class $\mathcal{A}(\bar{\sigma}, \varkappa)$. Let us consider a symmetric stable Lévy model with a nonzero diffusion part ($\sigma > 0$)

$$\psi(u) = \mathrm{i}\mu u - \sigma^2u^2/2 + \vartheta(u), \quad \vartheta(u) = -\eta|u|^\alpha,\ 0<\alpha\le 1, \quad u\in\mathbb{R}.$$

For any δ satisfying $0 < \delta < \alpha$ and $M > 0$ define

$$\psi_\delta(u) = \mathrm{i}\mu u - \sigma^2u^2/2 + \vartheta_\delta(u),$$

with

$$\vartheta_\delta(u) = -\eta|u|^\alpha\mathbf{1}_{\{|u|\le M\}} - \eta M^\delta|u|^{\alpha-\delta}\mathbf{1}_{\{|u|>M\}}.$$

Then $\phi_\delta(u) = \exp(\psi_\delta(u))$ is the characteristic function of some Lévy process and

$$\phi_\delta(u) = \phi(u), \quad |u| \le M,$$

where $\phi(u) = \exp(\psi(u))$. Indeed, $\vartheta_\delta(u)$ is a continuous, non-positive, symmetric function which is convex on $\mathbb{R}_+$ for large enough M. According to the well known Pólya criteria (see e.g. [22]), the function $\exp(\xi\vartheta_\delta(u))$ is the c. f. of some absolutely continuous distribution for any $\xi > 0$. In particular, for any natural k the function $\exp(\vartheta_\delta(u)/k)$ is the c. f. of some absolutely continuous distribution. Hence, $\exp(\vartheta_\delta(u))$ is the c.f. of some infinitely divisible distribution. Define now two affine (in fact, Lévy) models θ_1 and θ_2 corresponding to the c.f. characteristic exponents ψ and ψ_δ, respectively. Let ν_{θ_1} and ν_{θ_2} be the corresponding Lévy measures. It holds

$$\chi^2(P_{\theta_1}^{\otimes n}, P_{\theta_2}^{\otimes n}) = n\chi^2(p_{\theta_1}, p_{\theta_2}) = n\int_{\mathbb{R}}\frac{|p_{\theta_1}(y)-p_{\theta_2}(y)|^2}{p_{\theta_1}(y)}\,\mathrm{d}y,$$

where p_{θ_1} and p_{θ_2} are densities corresponding to c.f. ϕ_{θ_1} and ϕ_{θ_2} respectively. Using the asymptotic inequality

$$p_{\theta_1}(y) \gtrsim |y|^{-(\alpha+1)}, \quad |y| \to \infty$$

and the fact that the density of the stable law does not vanish on any compact set in $\mathbb{R}$, we derive

$$n\chi^2(p_{\theta_1}, p_{\theta_2}) \le nC_1 \int_{|y|\le A} |p_{\theta_1}(y) - p_{\theta_2}(y)|^2 \,\mathrm{d}y + nC_2 \int_{|y|>A} |y|^{\alpha+1} |p_{\theta_1}(y) - p_{\theta_2}(y)|^2 \,\mathrm{d}y = nC_1 I_1 + nC_2 I_2$$

for large enough $A > 0$ and some constants $C_1, C_2 > 0$. Parseval's identity implies

$$\begin{aligned} I_1 &\le \frac{1}{2\pi} \int_{\mathbb{R}} |\phi_{\theta_1}(u) - \phi_{\theta_2}(u)|^2 \,\mathrm{d}u \\ &\le \frac{1}{2\pi} \int_{|u|>M} \mathrm{e}^{-\sigma^2|u|^2} \,\mathrm{d}u \lesssim M^{-1} \mathrm{e}^{-\sigma^2 M^2}, \quad M \to \infty, \\ I_2 &\le \frac{1}{2\pi} \int_{|u|>M} |(\phi_{\theta_1}(u) - \phi_{\theta_2}(u))'|^2 \,\mathrm{d}u \\ &\lesssim \int_{|u|>M} |u|^2 \mathrm{e}^{-\sigma^2|u|^2} \,\mathrm{d}u \lesssim M \mathrm{e}^{-\sigma^2 M^2}, \quad M \to \infty. \end{aligned}$$

The choice $M \asymp [\sigma^{-2} \log(n \log^\beta n)]^{1/2}$ with some $\beta > 0$ yields

$$n\chi^2(p_{\theta_1}, p_{\theta_2}) < 1/2$$

for large enough n. On the other hand,

$$\bar{\vartheta}(u) - \bar{\vartheta}_\delta(u) = -\eta \int_{-1}^{1} [|u|^\alpha \mathbf{1}_{\{|u|>M\}} - |u+w|^\alpha \mathbf{1}_{\{|u+w|>M\}}] \,\mathrm{d}w + \eta M^\delta \int_{-1}^{1} [|u|^{\alpha-\delta} \mathbf{1}_{\{|u|>M\}} - |u+w|^{\alpha-\delta} \mathbf{1}_{\{|u+w|>M\}}] \,\mathrm{d}w$$

with

$$\bar{\vartheta}(u) := \int_{-1}^{1} [\vartheta(u) - \vartheta(u+w)] \,\mathrm{d}w,$$

$$\bar{\vartheta}_\delta(u) := \int_{-1}^{1} [\vartheta_\delta(u) - \vartheta_\delta(u+w)] \,\mathrm{d}w.$$

Using the identity

$$\int_{-1}^{1} [|u|^\alpha \mathbf{1}_{\{|u|>M\}} - |u+w|^\alpha \mathbf{1}_{\{|u+w|>M\}}] \,\mathrm{d}w = |u|^\alpha \int_{-1}^{1} [1 - |1 + w/u|^\alpha] \,\mathrm{d}w = 2 \sum_{k=1}^{\infty} \binom{\alpha}{2k} \frac{|u|^{\alpha-2k}}{2k+1}, \tag{40}$$

which holds for any $|u| > M + 1$ and $M > 1$, we get

$$\left| \int_{\mathbb{R}} [\bar{\vartheta}(u) - \bar{\vartheta}_\delta(u)] \,\mathrm{d}u \right| \,\mathrm{d}u \gtrsim M^{\alpha-1}, \quad M \to \infty$$

for any $0 < \alpha < 1$. Denote

$$\rho_{\theta_1}(x) := \int_{\mathbb{R}} e^{iux}\bar{\vartheta}(u)\,du = \left(1 - \frac{\sin x}{x}\right)\nu_{\theta_1}(x),$$
$$\rho_{\theta_2}(x) := \int_{\mathbb{R}} e^{iux}\bar{\vartheta}_\delta(u)\,du = \left(1 - \frac{\sin x}{x}\right)\nu_{\theta_2}(x),$$

then the Fourier inversion formula implies

$$\begin{aligned} d(\theta_1,\theta_2) &:= \sup_{x\in\mathbb{R}} |\rho_{\theta_1}(x) - \rho_{\theta_2}(x)| \\ &\geq \left|\int_{\mathbb{R}} [\bar{\vartheta}(u) - \bar{\vartheta}_\delta(u)]\,du\right| \gtrsim M^{\alpha-1}, \quad M \to \infty. \end{aligned} \tag{41}$$

Asymptotic expansion (40) shows that there is a constant C depending on η such that

$$|u|^{2-\alpha}\bar{\vartheta}(u) \leq C, \qquad |u|^{2-\alpha}\bar{\vartheta}_\delta(u) \leq C, \quad u \in \mathbb{R}.$$

Hence, taking $\bar{\sigma}^2 = \sigma^2/2$, $\varkappa = 2 - \alpha$, we conclude that both models θ_1 and θ_2 are in the class $\mathcal{A}(\bar{\sigma},\varkappa)$.

Acknowledgements

This work has been supported in part by the Collaborative Research Center 'Economic Risk' (SFB 649) of the German Research Foundation (DFG).

Appendix

A.1. Regularity properties of affine processes

The next lemma provides bounds on the growth of the derivatives of the conditional characteristic function of an affine process.

Lemma A.1. *If for some natural $k > 0$, the Lévy measure $\nu^{(0)}$ satisfies*

$$\int_{\{\|x\|>1\}} \|x\|^k \nu^{(0)}(dx) < \infty, \tag{42}$$

then functions $\psi_0(u,s)$ and $\psi_1(u,s)$ from the representation (1) are in $C^{k+1}(\mathbb{R}_+)$ as functions of s. Moreover, for any fixed $x \in \mathcal{D}$ and $s \in [0,T]$ the following estimates hold

$$\left|\frac{\partial^{\mathbf{m}+j}\phi(u|s,x)}{\partial x^{\mathbf{m}}\partial s^j}\right| \leq C\|u\|_2^{2(|\mathbf{m}|+j)}, \quad \|u\|_2 > 1, \quad j = 0,\ldots,k+1, \tag{43}$$

where $\mathbf{m}$ is a multi index and C is a positive constant depending on s and x.

Proof. The existence of derivatives in s up to order $k+1$ was proved in [8, Lemma 6.5]. As was also shown in [8, Section 7], for any fixed s and x, the function $\phi(u|s,x)$ is a c.f. of some infinitely divisible distribution, implying that

$$\log\phi(u|s,x) = -u^\top Au + i(b,u) + c + \int_{\mathcal{D}\setminus\{0\}} (e^{iy^\top u} - 1 - (\chi(y),u))M(dy),$$

where $A \in \mathbb{R}^{d\times d}$, $b \in \mathbb{R}^d$, $c \in \mathbb{R}$ and M is non-negative Borel measure on $\mathcal{D} \setminus \{0\}$. It remains to note that A, b, c and M depend linearly on x and are smooth in s, provided the condition (42) holds. $\square$

References

[1] Y. Aït-Sahalia, J. Jacod, Estimating the degree of activity of jumps in high frequency financial data, Annals of Statistics 37 (5A) (2009) 2202–2244.
[2] E. Andreou, E. Ghysels, A. Kourtellos, Forecasting with mixed-frequency data, in: Michael P. Clements, David F. Hendry (Eds.), Oxford Handbook on Economic Forecasting, 2010.
[3] I.V. Basawa, P.J. Brockwell, Nonparametric estimation for nondecreasing Lévy processes, Journal of the Royal Statistical Society. Series B 44 (1982) 262–269.
[4] D. Bates, Post-'87 crash fears in the S&P 500 futures option market, Journal of Econometrics 94 (2000) 181–238.
[5] D. Bates, Maximum likelihood estimation of latent affine processes, Review of Financial Studies (2005) 909–965.
[6] D. Belomestny, Spectral estimation of the fractional order of a Lévy process, Annals of Statistics 38 (1) (2009) 317–351.
[7] R. Cont, C. Mancini, Nonparametric tests for analyzing the fine structure of price fluctuations, SSRN Paper, 2004.
[8] D. Duffie, D. Filipović, W. Schachermayer, Affine processes and applications in finance, The Annals of Applied Probability 13 (2003) 984–1053.
[9] D. Duffie, J. Pan, K. Singleton, Transform analysis and asset pricing for affine jump diffusions, Econometrica 68 (2000) 1343–1376.
[10] J.E. Figueroa-López, Nonparametric estimation of Lévy processes with a view towards mathematical finance, Ph.D. Thesis, Georgia Institute of Technology, 2004. http://etd.gatech.edu.No.etd-04072004-122020.
[11] J.E. Figueroa-Lopez, Nonparametric estimation of time-changed Lévy models under high-frequency data, Advances in Applied Probability 41 (4) (2009) 1161–1188.
[12] G. Jongbloed, F.H. van der Meulen, A.W. van der Vaart, Nonparametric inference for Lévy-driven Ornstein–Uhlenbeck processes, Bernoulli 11 (5) (2005) 759–791.
[13] M. Keller-Ressel, W. Schachermayer, J. Teichmann, Affine processes are regular, Probability Theory and Related Fields, 2008 (in press).
[14] H. Masuda, Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps, Stochastic Processes and their Applications 117 (1) (2007) 35–56.
[15] F. Merlevéde, M. Peligrad, E. Rio, Bernstein inequality and moderate deviation under strong mixing conditions, Working Paper, 2009.
[16] M. Neumann, M. Reiß, Nonparametric estimation for Lévy processes from low-frequency observations, Bernoulli 15 (1) (2007) 223–248.
[17] H. Rubin, H.G. Tucker, Estimating the parameters of a differential process, Annals of Mathematical Statistics 30 (1959) 641–658.
[18] K. Singleton, Estimation of affine asset pricing models using the empirical characteristic function, Journal of Econometrics 10 (2001) 111–141.
[19] C.J. Stone, Optimal global rates of convergence for nonparametric regression, Annals of Statistics 10 (1982) 1040–1053.
[20] M. Tao, Y. Wang, Q. Yao, J. Zou, Large volatility matrix inference via combining low-frequency and high-frequency approaches, Technical Report, 2010.
[21] A. Tsybakov, Introduction to Nonparametric Estimation, in: Springer Series in Statistics, Springer, 2008.
[22] N. Ushakov, Selected Topics in Characteristic Functions, in: Modern Probability and Statistics, VSP, Utrecht, 1999.
[23] R. Yokoyama, Moment bounds for stationary mixing sequences, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 52 (1980) 45–57.